

Maier, Wolfgang

Diss. Mainz QU





Digitized by the Internet Archive  
in 2026

[https://archive.org/details/IA41555001\\_0022](https://archive.org/details/IA41555001_0022)



Aus der Psychiatrischen Klinik  
der  
Johannes-Gutenberg-Universität Mainz

Untersuchung zur Eignung  
der  
HAMILTON-Depressionsskala als Indikator des Schweregrads  
depressiver Syndrome

Inauguraldissertation  
zur Erlangung des Doktorgrades der Medizin  
der Johannes-Gutenberg-Universität Mainz  
dem Fachbereich Medizin  
vorgelegt

von  
Wolfgang Maier  
aus  
Hof/Saale

Mainz 1985



## Inhaltsverzeichnis

### 1. EINLEITUNG

### 2. FACHCHARAKTERISTIK UND WIRKUNG

### 3. INHALTSANALYSE UND BEWERTUNG

### 4. BEWERTUNG DER BEFUND-DEUTUNG

### 5. BEWERTUNG DER BEFUND-DEUTUNG

### 6. VERGLEICHUNG DER BEFUND-DEUTUNG

### 7. ZUSAMMENFASSUNG

### 8. LITERATUR

Dekan : Prof. Dr. Michaelis

1. Gutachter : Prof. Dr. Benkert

2. Gutachter : Prof. Dr. Rieger

Tag der Promotion : 19.11.1985



## Inhaltsverzeichnis

|                                                     |    |
|-----------------------------------------------------|----|
| 1. EINLEITUNG                                       | 1  |
| 2. FAKTORENANALYSE DER HAMD                         | 5  |
| 3. Homogenitätsanalysen der HAMD                    | 20 |
| 4. Reduktion der HAMILTON-Depressions-Skala         | 27 |
| 5. Weitere Reduktion der HAMILTON-Depressions-Skala | 34 |
| 6. Validierung der extrahierten Subskalen           | 42 |
| 7. Zusammenfassung                                  | 52 |
| 8. Literatur                                        | 54 |



## 1. EINLEITUNG

Die Hamilton-Depressions-Skala (HAMD) dient zur Beurteilung des Schweregrades des depressiven Syndroms bei Patienten, die als depressiv diagnostiziert wurden. Sie wurde 1960 in der Version mit 17 Items entwickelt (19) und 1967 um 4 Items erweitert (23). Seither wird die HAMD in der überwiegenden Zahl der Evaluationsstudien zum Wirksamkeitsnachweis von Antidepressiva verwendet. Sie stellt somit den Vergleichsmaßstab zwischen verschiedenen Antidepressivastudien dar.

Placebokontrollierte Evaluationsstudien zur Untersuchung der antidepressiven Wirksamkeit von Standardantidepressiva erbrachten keine durchgängige signifikante Überlegenheit gegenüber Placebo (6). Das widerspricht der klinischen Erfahrung, die eine klare und durchgängige Überlegenheit der Standardantidepressiva nahelegt (7). Als eine mögliche Erklärung für diese Diskrepanz müssen Mängel des verwendeten Beurteilungsinstruments für den antidepressiven Effekt diskutiert werden. Eine Untersuchung möglicher Mängel von Beurteilungsinstrumenten für den antidepressiven Effekt muß sich vor allem auf die HAMD stützen, da diese in der Regel für diesen Zweck verwendet wird.

Zwar weisen die Items in der HAMD überwiegend eine sehr geringe Reliabilität auf (9), (14); der Globalsummenscore zeigt jedoch in allen Studien überwiegend ausreichende Werte (jedenfalls größer 0,80: (19), (9), (14)).

Die Validität der HAMD ist neuerdings jedoch strittig (2). HAMILTON belegte die Validität der HAMD ausschließlich im Rahmen von Querschnittsuntersuchungen; der Schweregrad des depressiven Syndroms bzw. der depressiven Erkrankung wurde mit dem Summenwert der HAMD in korrelative Beziehung gesetzt. Der resultierende Wert liegt ausreichend hoch (größer als 0,80 in



(21)); dies konnte auch in anderen Studien bestätigt werden (26); allerdings werden auch geringere Werte angegeben (51).

Die Fähigkeit der HAMD, Veränderungen im Therapieverlauf valide abzubilden, wurde dagegen bisher noch nicht geprüft; sie wird allerdings von einigen Autoren in Zweifel gezogen (z.B. (36) ).

Im folgenden soll geprüft werden, ob der Summenscore der HAMD ein geeigneter Maßstab für den Schweregrad depressiver Symptomatik ist. Im Vordergrund steht aber nicht die externe Validität (Korrelation mit der Globalbeurteilung des depressiven Syndroms), sondern die Homogenität und die Populationsunabhängigkeit der Struktur der Skala. Homogenität bedeutet dabei, daß der Schweregrad des depressiven Syndroms bzw. der Summenscore der Schweregradskala in jedem einzelnen Item bzw. in Itemgruppen gleichermaßen repräsentiert wird; Items, die zum Schweregrad bzw. zum Summenscore (also der latenten Variablen) nicht in einem ausreichend engen Verhältnis stehen, tragen nur zur Erhöhung der Varianz bei. Eine Erhöhung der Varianz ist jedoch unerwünscht, da diese bei statistischen Tests nur den Fehler 2. Art erhöht. So wird die Gefahr vergrößert, daß ein tatsächlicher Unterschied zwischen Gruppen bezüglich des Schweregrades der Depression nicht als signifikanter Unterschied im Test nachgewiesen wird. Unter Populationsunabhängigkeit ist zu verstehen, daß die Skala in verschiedenen zusammengesetzten Populationen jeweils dieselbe Eigenschaft mißt. Die Populationsunabhängigkeit wurde vor allem von CATTELL (8) als relevantes Optimierungskriterium für Skalen hervorgehoben; die beiden Kriterien Homogenität, Populationsunabhängigkeit wurden von BECH (2) neben der Reliabilität als die relevantesten Optimierungskriterien für Schweregradskalen herausgearbeitet.

Mit den Methoden der Faktorenanalyse und der Homogenitätsanalyse kann geprüft werden, ob die Items eine Skala den Schwere-



grad einer latenten Variablen repräsentieren und ob die Struktur der Skala populationsunabhängig ist.

Dabei zerlegen Faktorenanalysen den korrelativen Zusammenhang zwischen allen Items der Skala in verschiedene lineare Komponenten (43); die Populationsunabhängigkeit der Struktur der Skala zeigt sich in der Stabilität der Lösungen der Faktorenanalysen in verschiedenen Populationen. Die Homogenitätsanalysen sind im Gegensatz zur Faktorenanalyse probabilistische Meßmodelle, die in jeder Ausprägung eines Items bei einem Patienten die Realisierung eines Zufallsprozesses sehen. Weiter Zusatzannahmen legen dann ein Wahrscheinlichkeitsmodell fest, in dem jeder Patient und jedes Item durch je einen Parameter gekennzeichnet sind. In diesen Modellen sind alle Items gleichermaßen Maßstab für den Schweregrad einer latenten Variablen; zugleich postulieren diese Modelle (allerdings in unterschiedlichem Ausmaß) die Populationsunabhängigkeit der Struktur der Skala. Modellanpassungstests prüfen, ob ein gegebener Datensatz diesen Modellannahmen entspricht (3), (16), (38).

In Abschnitt 2 werden daher Faktorenanalysen, in Abschnitt 3 Homogenitätsanalysen von HAMD-Daten aus verschiedenen Patientenstichproben durchgeführt, um zu prüfen, ob die HAMD ein geeigneter und stabiler Maßstab für den Schweregrad der Depression ist.

Die folgenden Abschnitte 4, 5, 6 befassen sich mit den Konsequenzen der Ergebnisse der Abschnitte 2 und 3.

Ein weiteres bislang unzureichend bearbeitetes Problem der Therapieevaluationsforschung besteht in der Herausarbeitung von näher bestimmten Indikationen (Prägnanztypen des depressiven Syndroms) für bestimmte Antidepressiva. Zwar werden manchmal Indikationen von bestimmten Antidepressiva bei bestimmten Prägnanztypen des depressiven Syndroms angegeben (z.B. (15)), doch sind diese praktischen Regeln weitgehend ungeprüft bzw. unzureichend reproduziert (z.B. (25)). Widersprüchliche



Ergebnisse treten auch bei der Evaluation der Therapie mit Adjuvantien auf (z.B. für die zusätzliche Verordnung von Diazepam bei ängstlich-depressiven Syndromen (44) im Vergleich zu (41)).

Ein möglicher Grund hierfür ist der Mangel an geeigneten Beurteilungsinstrumenten zur Erfassung des Schweregrades von Prägnanztypen des depressiven Syndroms. In einigen Studien wurden Items oder Subskalen der HAMD zur Beschreibung des Schweregrades von Prägnanztypen des depressiven Syndroms verwendet. Die Verwendung von Items ist dabei wegen der zu geringen Reliabilität für die Mehrzahl der Items (z.B. (14)), nur möglich, wenn sehr große Fehlermöglichkeiten in Rechnung gestellt werden. Daher erscheint zunächst die Verwendung von Subskalen angemessener. Doch wurde die Verwendbarkeit von Subskalen der HAMD für die Beurteilung des Schweregrads depressiver Subtypen bisher nicht geprüft. Da die Struktur der HAMD zur Diskussion steht, kann die Verwendbarkeit von Subskalen zu Evaluationszwecken untersucht werden. Die in der Evaluationsforschung benutzten Subskalen beschreiben nämlich in der Regel Faktoren, die nur auf Items relevante Ladungen tragen, die die Symptome eines klinischen Syndroms beschreiben. So kann in Abschnitt 2 (Faktorenanalyse) geprüft werden, ob in der HAMD Subskalen extrahierbar sind, die Prägnanztypen des depressiven Syndroms beschreiben.



## 2. FAKTORENANALYSE DER HAMD

Zu diesen faktorenanalytischen Fragestellungen sagen die bisherigen Analysen (15, 19, 20, 23, 33, 37, 39, 47) folgendes aus:

1. Ein Faktor, der auf allen Items bzw. auf fast allen mittleren oder hohen Ladungen trägt, konnte nicht in allen Faktorenanalysen extrahiert werden; ein Generalfaktor (43, Seite 58) ist aber wünschenswert, wenn die HAMD in ihrem Globalsummenskore den Schweregrad der depressiven Symptomatik repräsentieren soll.
2. Faktoren, die auf einigen Items mittlere bzw. hohe Ladungen tragen, auf den anderen Items dagegen niedrige Ladungen aufweisen, zeigen in den verschiedenen Analysen teilweise ungenügende Gemeinsamkeiten. So hat HAMILTON (20) einen bipolaren Faktor isoliert, auf dem die Items "Hemmung" bzw. "Erregung" Koeffizienten mit hohen Absolutwerten und jeweils entgegengesetztem Vorzeichen aufweisen; solche bipolaren Faktoren konnten in anderen Untersuchungen - z.B. (1) - nicht reproduziert werden.

Auf dem Hintergrund dieser widersprüchlichen Resultate haben MÖLLER und v. ZERSEN (34) für den deutschsprachigen Raum verbindliche faktorenanalytische Zerlegungen der HAMD gefordert. Hierzu sind Analysen aus verschiedenen Zentren notwendig.

Die bisher einzige faktorenanalytische Zerlegung der deutschsprachigen Version der HAMD (1) wurde an einer nach der ICD klassifizierten Patientenstichprobe durchgeführt, in der sich Patienten mit neurotischer und endogener Depression sowie mit schizoaffektiver Psychose befanden; ein Fünftel der Patienten war älter als 60 Jahre. Dabei resultierte als einzige Lösung, die sowohl im Verlauf als auch bei Zufallsteilung der Population hinreichende Stabilität aufwies, eine Einfaktorenlösung;



Eine mehrfaktorielle Auswertung der HAMD bzw. eine Beurteilung von Prägnanztypen der depressiven Symptomatik mit der HAMD ist damit in Frage gestellt. Außerdem stellt der extrahierte Faktor keinen Generalfaktor dar; mehr als ein Fünftel der Items trägt Ladungskoeffizienten kleiner als 0,4. Damit ist jedoch auch die Eignung der HAMD zur Beurteilung des Schweregrades der depressiven Symptomatik zweifelhaft. Angesichts der Verbreitung der HAMD in der Evaluationsforschung ist dieses Resultat beunruhigend. In (1) wird nicht weiter diskutiert, wodurch dieses Ergebnis bedingt ist; insbesondere wird nicht untersucht, ob die Zusammensetzung der Stichprobe hierfür verantwortlich ist. Es sind also Wiederholungsstudien notwendig, in denen untersucht wird, ob die Resultate in (1) durch die Selektion der Patienten bedingt sind.

## 2.1 Untersuchungsmethoden

Charakterisierung der Stichprobe (s. Tab. 1):

Stichprobe 1 und Stichprobe 3 umfassen dieselben Patientengruppe; die beiden Stichproben unterscheiden sich bezüglich verschiedener Erhebungszeitpunkte der HAMD. Stichprobe 2 umfaßt ebenso wie Stichprobe 1 stationäre Patienten, die innerhalb der ersten vier Tage des stationären Aufenthaltes nach der HAMD beurteilt wurden; Stichprobe 2 ist dabei heterogener zusammengesetzt (bezüglich der diagnostischen Zuordnung) als Stichprobe P1.

Stichprobe P1: Sämtliche Patienten erfüllen die Kriterien von Major Depressive Disorder (endogenous type - sicher) nach RDC (31); es mußte mindestens die 2. abgesetzte depressive Phase vorliegen. Diese Stichprobe umfaßt sämtliche stationäre Patienten, die an einer von fünf Antidepressivastudien teilnahmen. Die Beurteilung auf der HAMD erfolgte nach einer mindestens viertägigen antidepressivafreien Periode vor Ansetzen der antidepressiven Medikation.



Stichprobe P2: Diese Stichprobe besteht aus allen Patienten, die in einem festen Zeitraum in die Psychiatrische Universitätsklinik Mainz aufgenommen wurden und sich einer mindestens vierzehntägigen stationären Therapie unterzogen und den Kriterien A und B von Minor Depressive Disorder gemäß RDC genügten; Patienten, die nach RDC an einer Schizophrenie litten, wurden (ebenso wie in P1) ausgeschlossen.

Die Stichproben P1 und P2 sind disjunkt.

Stichprobe P3: Die Patienten dieser Gruppe sind sämtliche Patienten, die auch der Stichprobe P1 angehören; sie wurden am Ende der Antidepressivastudien (nach 21 Tagen) erneut mit der HAMD beurteilt. Bei 19 Patienten in P1 mußten die Antidepressivastudien abgebrochen werden; diese Patienten gehören nicht der Stichprobe P3 an.

In keiner der Stichproben wurden Patienten aufgenommen, die an einer ernsthaften körperlichen Erkrankung, einem Anfallsleiden oder einem Abusus von Medikamenten, Alkohol oder Drogen litten.

#### 2.1.1 Durchführung der Untersuchung:

Trainierte Beurteiler wendeten in sämtlichen drei Stichproben die deutsche Version der HAMD mit 21 Items an (Version: CIPS (10)). In der Stichprobe P2 wurde die Exploration durch ein strukturiertes Interview nach present state examination (PSE) (52) eingeleitet. Die Beurteiler der Stichproben P1 und P3 beteiligten sich nicht am Rating in der Stichprobe P2, die Ratings fanden jeweils vormittags statt.

#### 2.1.2 Analysenmethoden:

Jeder der drei gewonnenen Datensätze wurden einer Itemanalyse



(Mittelwerte, Varianzen, Trennschärfen, innere Konsistenz) einer Faktorenanalyse (Hauptfaktorenanalyse mit schiefwinkliger und orthogonaler Rotation) unterzogen. In den Faktorenanalysen wurden sämtliche Faktoren berücksichtigt, die Eigenwerte größer 1 aufweisen (40, S. 125).

Tabelle 1

Zusammensetzung der untersuchten Patientengruppe

| Anzahl der Patienten                                                | P1<br>n = 107 | P2<br>n = 98 | P3<br>n = 88 |
|---------------------------------------------------------------------|---------------|--------------|--------------|
| Geschlecht                                                          |               |              |              |
| weiblich                                                            | n = 69        | n = 71       | n = 62       |
| männlich                                                            | n = 38        | n = 27       | n = 26       |
| Alter<br>(Mittelwert)<br>(jedenfalls 20 - 60 J.)                    | 42,1 J        | 40,2 J.      | 41,0 J       |
| Major Depressive Disorder<br>(i.S. RDC)                             | alle          | n = 43       | alle         |
| Major Depressive Disorder<br>endogenous type<br>(sicher) (i.S. RDC) | alle          | n = 22       | alle         |
| bipolar<br>(bipolar I oder II)                                      | n = 33        | n = 7        | n = 26       |
| unipolar Major Depressive Disorder                                  | n = 74        | n = 36       | n = 62       |
| Minor Depressive Disorder (i.S. RDC)                                | 0             | 35           | 0            |
| Schizoaffective Disorder<br>(depressive type)<br>(i.S. RDC)         | 0             | 20           | 0            |



## 2.2 Ergebnisse

### 1. Mittelwerte, Itemanalysen:

- a) Der Mittelwert von P1 (22,26;  $s = 6,70$ ) und P3 (12,46;  $s = 7,73$ ) liegt unter den in (1) angegebenen Werten: 24,19 ( $s = 7,32$ ) für Tag 0 und 14,18 ( $s = 8,86$ ) für Tag 21 (jeweils 17-Items-Version). In der Stichprobe von HAMILTON (11) wird ein Mittelwert von 23,03 angegeben. Die Vergleichsstichproben sind dabei jeweils nach anderen als den hier verwendeten Kriterien erhoben.
- b) Trennschärfe: Die Trennschärfe der Items Nr. 14 (Libido-verlust) (0,05), Nr. 15 (Hypochondrie) (0,04), Nr. 16 (Gewichtsverlust) (0,03) weisen in mindestens einer Stichprobe Werte auf, die nicht signifikant (Niveau 1 %) von 0 abweichen (HAMD-Version mit 17 Items)\*. Entsprechend sind in der Version mit 21 Items die Items Nr. 19 (Depersonalisation) (0,01) und 21 (Zwangssymptome) (0,03) in mindestens einer der untersuchten Stichproben nicht hinreichend trennscharf (d.i. keine signifikante Abweichung von Null - Niveau 1 %).
- c) Die Konsistenzkoeffizienten (innere Konsistenz nach CRONBACH) liegt im gewünschten Bereich ((29) S. 309). Die Konsistenz liegt bei der Version mit 21 Items geringfügig höher als in der Version mit 17 Items. Wird der Konsistenzkoeffizient als Reliabilitätsschätzung verwendet, so zeigt er eine ausreichende Reliabilität der Skala an.

\* Es sind jeweils die niedrigsten Werte angegeben.



Tabelle 2

Statistische Kennwerte

|                                                    | P1    | P2    | P3    |
|----------------------------------------------------|-------|-------|-------|
| HAMD-Version mit 17 Items:                         |       |       |       |
| Mittelwert                                         | 22,26 | 18,19 | 12,46 |
| Standardabweichungen                               | 6,70  | 7,34  | 7,73  |
| in Globalsumme<br>repräsentierter<br>Varianzanteil | 22,1% | 24,2% | 33,1% |
| Konsistenz<br>(Alpha-Koeffizient<br>v. CRONBACH)   | 0,72  | 0,73  | 0,83  |
| HAMD-Version mit 21 Items:                         |       |       |       |
| Mittelwert                                         | 24,98 | 20,83 | 14,82 |
| Standardabweichungen                               | 6,95  | 7,69  | 8,05  |
| in Globalsumme<br>repräsentierter<br>Varianzanteil | 21,5% | 23,6% | 32,5% |
| (Alpha-Koeffizient<br>v. CRONBACH)                 | 0,73  | 0,75  | 0,85  |

2. Extrahierbarkeit von Faktoren zur Beurteilung von Subtypen des depressiven Syndroms:

Bei der Darstellung verschiedener Qualitäten in einer Skala bedient man sich in der Literatur vor allem zweier Rotationsprinzipien: der orthogonalen und der schiefwinkligen Rotation.

In sämtlichen untersuchten Stichproben ergeben sich für die Version mit 17 Items jeweils 4 Faktoren; diese Lösung ist



Tabelle 3

Faktorenanalysen für Stichproben P1, P2, P3 für HAMD (17 Items).  
Schiefwinkelige Rotation; Items mit Ladungen kleiner 0,40 sind nicht aufgeführt;  
Ladungen der Items auf den Faktoren sind in Klammern angegeben.

| Stichproben                                                                                                               | Faktoren | Faktor<br>"Hemmung"                                                                                                              | Faktor<br>"Erregung -<br>psychische Angst"                                                                            | Faktor<br>"Hypochondrie-<br>somatische Angst"                                         | Faktor<br>"Schlafstörungen"                                                                       |
|---------------------------------------------------------------------------------------------------------------------------|----------|----------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|
| P1<br>Major Depressive Disorder<br>(endogenous)<br>akut<br>Wash-out-Phase<br>n=107                                        |          | depressive Stimmung<br>(.44)<br>Suizidalität (.79)<br>Hemmung (.69)<br>körperliche Symp-<br>-gastroint. (.44)<br>(1)             | depressive Stimmung (.40)<br>Erregung (.61)<br>Angst psychisch (.61)<br>Krankheitsuneinsichtig-<br>keit (.53)<br>(2)  | Arbeit (.83)<br>körperl. Symptome<br>-allgemein (.46)<br>Hypochondrie (.40)<br>(3)    | Einschlafstörungen(.61)<br>Durchschlafstörungen(.80)<br>Schlafstörungen am<br>Morgen (.60)<br>(4) |
| P2<br>Major u. Minor<br>Depressive Disorder<br>akut<br>teilweise unter Medikation<br>n=98                                 |          | depressive Stimmung<br>(.51)<br>Arbeit (.57)<br>Hemmung (.81)<br>körperl. Symp-<br>-allgem. (.44)<br>Genitalstörung (.46)<br>(1) | Suizidalität (.52)<br>Erregung (.52)<br>Angst psychisch (.72)<br>(2)                                                  | som. Angst. (.55)<br>körperl. Symptome<br>-allgem. (.55)<br>Hypochondrie (.45)<br>(4) | Einschlafstörungen(.69)<br>Durchschlafstörungen(.90)<br>Schlafstörungen am<br>Morgen (.63)<br>(3) |
| P3<br>Major Depressive<br>Disorder (endogenous)<br>nach 3 Wochen Therapie<br>unter standardisierter<br>Medikation<br>n=88 |          | Arbeit (.68)<br>Hemmung (.72)<br>körperl. Symptome<br>-allgem. (.40)<br>Gewichtsverlust(.45)<br>(2)                              | depressive Stimmung (.49)<br>Schuldgefühle (.45)<br>Suizidalität (.50)<br>Erregung (.59)<br>Angst psych. (.56)<br>(1) | somatische Angst (.49)<br>Hypochondrie (.70)<br>(4)                                   | Einschlafstörungen(.57)<br>Durchschlafstörungen(.70)<br>Schlafstörungen am<br>Morgen (.85)<br>(3) |

(.): bezeichnet die Rangfolge der Faktoren bezüglich des Anteils an erklärter Varianz



mit dem Scree-Graph-Kriterium (40, S. 127) verträglich. Bei Verwendung beider Rotationsprinzipien können diese Faktoren nach Leitsymptomen charakterisiert werden (Hemmung, Erregung - psychische Angst, Hypochondrie - somatische Angst, Schlafstörungen).

Die in Tabelle 3 gegebene Zuordnung der Items zu den Faktoren ist unabhängig von der verwendeten Rotationsmethode. Allerdings zeigt sich eine mangelnde Kongruenz der einzelnen Faktoren zwischen den verschiedenen Stichproben: Das Item "Suizidalität" lädt in der Stichprobe P1 auf dem durch das Leitsymptom "Hemmung" charakterisierten Faktor relevant (größer 0,4); in den beiden anderen Stichproben lädt "Suizidalität" auf dem Faktor, der durch "Erregung - psychische Angst" als Leitsymptom gekennzeichnet ist. Es entspricht der klinischen Erfahrung, daß im Therapieverlauf eine Erhöhung des Antriebspotentials häufig mit einem Weiterbestehen der depressiven Stimmung und der kognitiven Fehleinstellung (z.B. Suizidalität) einhergehen kann. Allein die "Schlafstörungen" weisen eine weitgehende Kongruenz auf.

Entsprechend zeigen die Ähnlichkeitskoeffizienten für den Globalvergleich zwischen den Faktorenlösungen verschiedenen Umfangs (Koeffizienten nach FISCHER und ROPPERT (17)) und die Ähnlichkeitskoeffizienten für den Vergleich der einzelnen Faktoren in der 4-Faktorenlösung (Koeffizienten nach BURT (45)) relativ niedrige Werte. Die FISCHER-ROPPERT Ähnlichkeitskoeffizienten zeigen für die resultierende 4-Faktoren-Lösungen beim Vergleich zwischen den verschiedenen Stichproben P1 und P2 bzw. P2 und P3 unzureichende Werte (jeweils kleiner 0,80), wenn als Normwerte die in (18) angegebenen Werte zugrunde gelegt werden (Tabelle 4). Die Faktorenkongruenzkoeffizienten nach BURT sind nur beim Vergleich der Faktoren, die durch das Leitsymptom "Schlafstörungen" charakterisiert sind, ausreichend hoch (Zugrundelegen der in (45) angegebene Normwert 0,90) (Tabelle 5).



Tabelle 4

Ähnlichkeitskoeffizienten von Hauptkomponentenlösungen  
(FISCHER-ROPPERT)

| Zahl der<br>vorkommenden<br>Faktoren<br><br>Stichprobenpaare | Faktoren |      |      |      |
|--------------------------------------------------------------|----------|------|------|------|
|                                                              | 1        | 2    | 3    | 4    |
| P1 / P2                                                      | 0,91     | 0,79 | 0,67 | 0,75 |
| P1 / P3                                                      | 0,87     | 0,82 | 0,88 | 0,92 |
| P2 / P3                                                      | 0,87     | 0,76 | 0,68 | 0,78 |

Tabelle 5

Faktorkongruenz-Koeffizient nach BURT für den Vergleich von  
Faktoren mit gleichen Leitsymptomen von jeweils zwei  
Stichproben

| Faktoren<br><br>Stich-<br>proben | Faktor    |            |                   |                |
|----------------------------------|-----------|------------|-------------------|----------------|
|                                  | "Hemmung" | "Erregung" | "Schlafstörungen" | "Hypochondrie" |
| P1 / P2                          | 0,68      | 0,64       | 0,93              | 0,20           |
| P1 / P3                          | 0,45      | 0,76       | 0,87              | 0,47           |
| P2 / P3                          | 0,75      | 0,67       | 0,88              | 0,72           |

Werte unter 0,8 zeigen eine unzureichende Vergleichbarkeit der  
einzelnen Faktoren in den verglichenen Populationen an.

Bei der Analyse der Version mit 21 Items ergeben sich in  
den verschiedenen Stichproben Lösungen verschiedenen Umfan-  
ges: Eine 5-Faktoren-Lösung in P3, eine 6-Faktoren-Lösung  
in P2 und eine 7-Faktorenlösung in P1. Die Lösungen sind  
jeweils mit dem Scree-Graph-Kriterium verträglich. Die  
Ähnlichkeitskoeffizienten für die 4-, 5-, 6- und 7-Fakto-



renlösungen sind nicht ausreichend (gemäß den in (18) angegebenen Normwerten), allerdings resultiert auch bei dieser Version in allen Stichproben ein Faktor mit dem Leitsymptom "Schlafstörungen", der eine ausreichende Stichprobenunabhängigkeit aufweist.

### 3. Extrahierbarkeit eines Generalfaktors zur Beurteilung des Schweregrades der depressiven Symptomatik.

Der Schweregrad kann sich sowohl in Symptomen (Items) als auch in Symptomverbänden bzw. Syndromen (Faktoren) niederschlagen; entsprechend können bei der Faktorenanalyse, die zur Extraktion eines als Schweregradindikator interpretierbaren Faktors durchgeführt wird, in jeder Stichprobe zwei verschiedene Korrelationsmatritzen zugrunde gelegt werden:

- a) die Korrelationsmatrix der einzelnen Items der HAMD (Schweregrad als Faktor erster Ordnung):
- b) die Korrelationsmatrix für jene Faktoren, die aus dem Items extrahiert werden konnten und auf eine nichtorthogonale Struktur rotiert wurden (Schweregrad als Faktor 2. Ordnung).

Für die Version mit 17 Items ergeben sich bei der Analyse der Korrelationsmatrix der einzelnen Items die folgenden Resultate (siehe Tabelle 6):

- a) Erwartungsgemäß hat in allen untersuchten Stichproben der unrotierte und varianzstärkste Faktor die meisten Items mit mittlerer oder höherer Ladung (Koeffizienten 0,4). Diese Items kennzeichnen keinen Subtyp des depressiven Syndroms, so daß der varianzstärkste Faktor am ehesten als Schweregradindikator interpretiert werden kann. Aber die Zahl der Items mit relevanter Ladung ist unbefriedigend: in den Stichproben P1 und P2 sind 6 Items und in der Stichprobe P3 sind 5 Items mit Werten unter 0,4 geladen.



Tabelle 6

Unrotierte varianzstärkste Faktoren  
in verschiedenen Stichproben

| Items | P 1  | P 2  | P 3  |
|-------|------|------|------|
| 1     | 0,80 | 0,65 | 0,89 |
| 2     | 0,51 | 0,26 | 0,65 |
| 3     | 0,57 | 0,21 | 0,72 |
| 4     | 0,32 | 0,51 | 0,51 |
| 5     | 0,30 | 0,53 | 0,30 |
| 6     | 0,44 | 0,54 | 0,46 |
| 7     | 0,61 | 0,62 | 0,69 |
| 8     | 0,57 | 0,38 | 0,67 |
| 9     | 0,45 | 0,21 | 0,65 |
| 10    | 0,76 | 0,59 | 0,53 |
| 11    | 0,21 | 0,59 | 0,53 |
| 12    | 0,50 | 0,69 | 0,60 |
| 13    | 0,47 | 0,64 | 0,70 |
| 14    | 0,10 | 0,54 | 0,37 |
| 15    | 0,13 | 0,25 | 0,13 |
| 16    | 0,10 | 0,59 | 0,31 |
| 17    | 0,41 | 0,15 | 0,27 |

Ein Generalfaktor kann in der Faktorenanalyse eines Datensatzes unter Umständen aus methodischen Gründen nicht extrahiert werden: bei zu geringer Heterogenität der Stichprobe ((43) Seite 295) bzw. bei Fehler eines heterogenen Kontextes (24) kann ein tatsächlich vorhandener Generalfaktor unter Umständen nicht extrahiert werden. Doch weder eine Heterogenisierung der Stichprobe (Vereinigung der beiden Stichproben P1 und P2 zu einer neuen Stichprobe, deren HAMD-Daten einer Faktorenanalyse unterworfen werden) noch die Einbettung der 17-Items-Version der HAMD in einen heterogenen Kontext (Analyse der 21-Items-Version anstelle der Version mit 17 Items) führt zu einer Erhöhung der Anzahl der auf dem varianzstärksten Faktor relevant ladenden Items.

Für die unrotierten varianzstärksten Faktoren ist die Stichprobenunabhängigkeit gegeben (Tabelle 4): der niedrig-



ste Wert für den FISCHER-ROPPERT-Koeffizienten liegt bei 0,87 (Vergleich von P2 und P3). Derselbe Koeffizient ergibt beim Vergleich der Einfaktorenlösung, die von BAUMANN (1) für verschiedene Ratingtage angegeben werden, stets Werte größer als 0,85 und mittlere Werte von 0,91.

Der Zusammenhang zwischen dem Globalsummenscore der HAMD und den unrotierten varianzstärksten Faktoren in den untersuchten Stichproben ist eng: die SPEARMAN-Korrelation zwischen den beiden Variablen (Globalsummenscore der 17-Items-Version bzw. mit Ladungskoeffizienten gewichteter Summenscore des unrotierten varianzstärksten Faktors) beträgt in P1 0,96, in P2 0,95 und in P3 0,97. Der Ähnlichkeitskoeffizient nach BURT zwischen den beiden Variablen zeigt in sämtlichen 3 Stichproben Werte größer 0,91. Entsprechend können der Globalsummenscore der HAMD und die unrotierten varianzstärksten Faktoren als weitgehend identisch angesehen werden.

Die Version mit 21 Items zeigt eine noch geringere Tendenz zur Generalfaktorenlösung als die kürzere Version: die Anzahl der Items mit irrelevanten Ladungen (Koeffizienten kleiner 0,4) beträgt in P1 9, in P2 11 und in P3 6. Die Ähnlichkeitskoeffizienten zwischen diesen Faktoren erbringen niedrige Werte (der niedrigste Wert ist 0,76 bei Vergleich des Faktors von P1 mit dem entsprechenden Faktor von P2).

- b) Beim Versuch einen Faktor zweiter Ordnung als Index für den Schweregrad der depressiven Symptomatik zu gewinnen, wird eine schiefwinklige Mehrfaktorenlösung (Faktoren erster Ordnung) zugrunde gelegt (in allen 3 Stichproben die in Tabelle 3 dargestellten Lösungen). Bei dieser Faktorenanalyse zweiter Ordnung (17 Items-Version) ergibt sich in den Stichproben P1 und P3 eine Einfaktorenlösung und in der Stichprobe P2 eine 2-Faktorenlösung; bipolare Faktoren resultieren nicht. Ein Generalfaktor kann aber in den Stichproben P1 und P2 weder unter den unrotierten noch



unter den rotierten Faktoren ausfindet gemacht werden; lediglich in P3 resultiert ein Generalfaktor.

Für die HAMD (Version 21 Items) resultiert bei der Faktorenanalyse zweiter Ordnung in der Stichprobe P3 ein Generalfaktor.

Auch bei Zugrundelegung der mehrfaktoriellen Zerlegung der HAMD (Version mit 21 Items), die von RHOADES (39) vorgeschlagen wurde, kann in keiner der untersuchten Stichproben ein Faktor zweiter Ordnung isoliert werden, der auf sämtlichen Faktoren erster Ordnung (diese sind die Subskalen von RHOADES) relevante Ladungen aufweist.

### 2.3 Diskussion

Die Itemanalysen zeigen, daß die Items der HAMD in den Mittelwerten und Trennschärfen breit streuen. Die Trennschärfen für die Items "Hypochondrie", "Genitalstörungen", "Gewichtsverlust" genügen bei relativ geringen Mittelwerten nicht den in (29 S. 126ff.) angegebenen Kriterien. Die Koeffizienten für die innere Konsistenz - die auch als Reliabilitätskoeffizienten interpretiert werden können - sind für die beiden Versionen der Skala ausreichend (29 S. 242). Die Erweiterung von 17 auf 21 Items bringt nur einen bescheidenen Zuwachs an innerer Konsistenz. Sieht man in der Erweiterung vor allem ein Mittel zur Reliabilitätssteigerung, so kann auf die Erweiterung verzichtet werden. Diese Konsequenz wird auch durch die relativ geringe Stichprobenunabhängigkeit der Faktorenanalyse der Skala mit 21 Items nahegelegt. Dieses Resümee entspricht dem von BAUMANN (1). Die Einfaktorenlösung ist dagegen verlaufsstabil und stichprobenunabhängig; sie zeigt außerdem in allen untersuchten Stichproben eine hinreichende Ähnlichkeit mit den in (1) isolierten Einfaktorenlösungen. Eine einfaktorielle Auswertung der Skala mit 17 Items erscheint unter diesem Gesichtspunkt adäquat; trotzdem genügt die Einfaktorenlösung den hier verwendeten Kriterien zur Extraktion der Faktoren



nicht. Der Faktor der Einfaktorenlösung kann nur bedingt als Schweregradindex der depressiven Symptomatik interpretiert werden: in allen untersuchten Stichproben zeigen mindestens ein Viertel der Items auf diesen Faktoren irrelevante Ladungen. Die Summenbildung beim Faktor der Einfaktorenklösung und die Globalsummenbildung der HAMD repräsentieren inhaltlich dasselbe, denn sie sind hoch miteinander korreliert (größer 0,95). Kann also der Faktor der Einfaktorenlösung nur bedingt als Maßstab für den Schweregrad einer depressiven Symptomatik gelten, so ist dies auch für den Globalsummenscore der HAMD der Fall. Andere Möglichkeiten zur faktorenanalytischen Isolierung eines Faktors, dessen Summenscore als Schweregrad der depressiven Symptomatik interpretiert werden könnte, versagen ebenso (Konstruktion von Faktoren zweiter Ordnung, Faktorenanalyse von Summenscores vorgegebener HAMD-Subskalen, Heterogenisierung der Stichprobe); so kann zusammenfassend der Schluß gezogen werden:

1. Unter den geprüften Bedingungen kann in der HAMD nur ein einziger Faktor isoliert werden, der keinen Subtyp der depressiven Symptomatik beschreibt; dieser Faktor ist jedoch nur bedingt als Indikator für den Schweregrad der depressiven Symptomatik interpretierbar.
2. In der HAMD sind keine weiteren Informationen zu extrahieren, die eine hinreichende faktorielle Validität aufweisen.
3. Die HAMD ist eine unökonomische Skala, da jener Faktor, auf den allein eine Auswertung gestützt werden kann, maximal ein Drittel der Varianz erklärt.

In jeder Stichprobe ergaben sich für die 17-Items-Version jeweils 4-Faktoren-Lösungen. Für die Version mit 21 Items ergaben sich Lösungen, deren Faktorenanzahlen stichprobenabhängig sind. Die meisten bisher vorgelegten Analysen geben vergleichbare Faktorenanzahlen an. Insbesondere zur bisher ein-



zigen Analyse mit der deutschsprachigen Version (1) bestehen keine wesentlichen Widersprüche: dort werden 1- bis 4-, 5-, 6-Faktorenlösungen als mit dem Scree-Graph-Kriterium verträglich angesehen, wobei die Anzahl der Eigenwerte, die größer als 1 sind, vier bis sechs beträgt. Die vier Faktoren, die für die Version mit 17 Items in allen untersuchten Stichproben extrahiert werden konnten, zeigen keine bipolaren Strukturen; sie lassen sich mit Leitsymptomen charakterisieren (Hemmung; Erregung - psychische Angst, Hypochondrie - somatische Angst, Schlafstörungen), die gebräuchlichen klinischen Differenzierungen entsprechen.

Allerdings sind die gewonnenen Faktoren nicht zur Erfassung des Schweregrades von Sybtypen des depressiven Syndroms geeignet; denn sie sind nicht hinreichend verlaufsstabil und stichprobenunabhängig. Diese Unstabilität kann die inhaltliche Heterogenität der Mehrfaktorenlösungen, die bisher in der Literatur angegeben wurden, erklären.

In (49) konnten ebenfalls vier Faktoren extrahiert werden, die durch dieselben Leitsymptome charakterisiert werden konnten; dagegen wurden in (5) vier Faktoren verwendet, bei denen "Erregung" und "Angst" je einen eigenen Faktor bezeichnen. Auch die von RHOADES et al. (39) vorgeschlagene Subskalierung der HAMD (Version mit 21 Items) konnte nicht reproduziert werden.

Die HAMD wurde wegen der hohen inhaltlichen Heterogenität dem Überwiegen der Items, die körperlich betonte Symptome abdecken, und der ungenügenden Beurteilung von Veränderungen des Schweregrades kritisiert (3, 36). Um die damit gegebenen Nachteile für die Therapieevaluationsforschung zu vermeiden, wurden zur Fremdbeurteilung des Schweregrades der depressiven Symptomatik andere Skalen vorgeschlagen (3, 36) (Itemumfänge 11 bzw. 10). Die hier vorgetragene Kritik an der HAMD unterstreicht die Notwendigkeit zur Konstruktion und zur parallelen Verwendung anderer Skalen zur Beurteilung des Schweregrades des depressiven Syndroms.



### 3. HOMOGENITÄTSANALYSEN DER HAMD

Die Faktorenanalyse belegte, daß aus der HAMD keine Subskalen zur Beurteilung von Prägnanztypen des depressiven Syndroms extrahiert werden können. Für die Beurteilung der gesamten depressiven Symptomatik ist die HAMD dagegen wegen der Stabilität der Einfaktorenlösung besser geeignet; doch selbst das ist zweifelhaft, da kein Generalfaktor aus der HAMD extrahiert werden konnte, der den faktorenanalytischen Beleg für eine Schweregradskala darstellt.

Die Faktorenanalyse selbst ist in einer Hinsicht mangelhaft: sie stützt sich nämlich auf Korrelationskoeffizienten (Produkt-Moment-Korrelationskoeffizienten), die am Ideal linearer Variablenzusammenhänge orientiert ist; lineare Variablenzusammenhänge können aber für Variablen, die psychopathologische Tatbestände beschreiben, in der Regel nicht angenommen werden.\* Das Fehlen eines Generalfaktors kann daher auch als Reflex einer methodologischen Insuffizienz der Faktorenanalyse interpretiert werden. Es bedarf also einer weiteren Klärung, ob die HAMD als Schweregradindikator für depressive Symptomatik geeignet ist.

Daher sollen die faktorenanalytisch untersuchten Daten zusätzlich im Rahmen einer anderen statistischen Meßtheorie untersucht werden, die Hinweise auf die Verwendbarkeit von Skalen als Schweregradindikator geben kann. Hierzu wählen wir Modellvorstellungen, die das Konzept der Homogenität formalisieren. Dieses Konzept beschreibt die folgenden Bedingungen:

1. Jedes Item einer Skala soll dieselbe Qualität beurteilen, die auch im Summenskore dieser Skala beurteilt wird.
2. Die Summenskores der Skalen sollen in verschiedenen Stichproben dieselbe Qualität beurteilen.

\*Für das Item "Tagesschwankungen" ist ein kurvilinearere Zusammenhang mit dem Schweregrad des depressiven Syndroms anzunehmen.



CATTELL (8) und anschließend BECH et al. (2) haben diese Bedingungen für Schweregradskalen erstmals als notwendig postuliert. Sie fordern neben einer hinreichenden Reliabilität eine hohe Homogenität und eine ausreichende Populationsunabhängigkeit (Transferability).

Zur näheren Begriffsbestimmung von "Homogenität" und "Populationsunabhängigkeit":

1. Homogenität:

Jedes Item soll gleichermaßen Maßstab für den Schweregrad sein. Jedes Item soll daher einen monotonen Zusammenhang zwischen seiner Ausprägung und dem Schweregrad (bzw. dem Globalsummenskore der Skala) aufweisen.

2. Populationsunabhängigkeit - Transferability:

In jeder Stichprobe soll der Zusammenhang zwischen den Items gleich sein; es sollen also in jeder Stichprobe die Rangordnungen der Items bezüglich ihrer mittleren Ausprägung in verschiedenen Stichproben (Untergruppen von Stichproben) gleich sein; bestehen zwischen zwei verschiedenen Stichproben (Untergruppen von Stichproben) für ein Item Unterschiede in den mittleren Itemausprägungen, so sollen auch entsprechend Unterschiede in den mittleren Ausprägungen der anderen Items bestehen. Damit ist gewährleistet, daß jedes Item in jeder Stichprobe (Untergruppe) Vergleichbares mißt, und daß der Beitrag einer Itemausprägung zur Summenskore in den verschiedenen Stichproben vergleichbar ist.

Genauer:

- a) Die Rangordnung der mittleren Ausprägungen von Items soll in verschiedenen Teilgruppen depressiver Patienten stabil sein; wenn ein Item also in jener Untergruppe der Stichprobe, die von den Patienten mit geringerer Depressionstiefe gebildet wird, höhere durchschnittliche Ausprägungen zeigt als ein anderes Item, so soll dies auch in jener Untergrup-



pe gelten, die von den Patienten mit höherer Depressions-tiefe gebildet wird.

- b) Ein Zusammenhang zwischen der Rangordnung der Items bezüglich ihrer durchschnittlichen Ausprägung und der Selektion von Patientengruppen soll nicht bestehen; Selektionskriterien sind z.B. Alter, Geschlecht, Diagnose oder Zufalls Gesichtspunkte - ; wenn also ein Item in der Gruppe der älteren Patienten eine höhere durchschnittliche Ausprägung aufweist als ein anderes Item, so soll dies auch in der Gruppe der Patienten mit jüngerem Alter der Fall sein usw.

Diese beiden Bedingungen charakterisieren probabilistische Meßmodelle. Das Meßmodell von MOKKEN (35) und das Meßmodell von RASCH (36) sind Beispiele für solche Modelle. Das Modell von MOKKEN basiert auf dem Homogenitätskoeffizienten von LOEVINGER und formalisiert die beiden Bedingungen "Homogenität" und - allerdings nur bedingt - "Populationsunabhängigkeit - Transferability". Das RASCH-Modell formalisiert die Bedingungen "Homogenität" und "Populationsunabhängigkeit - Transferability" in vollem Umfang.

Im RASCH-Modell ist jedoch noch eine andere Bedingung formalisiert:

### 3. Suffizienz-Summierbarkeit:

Der Summenscore soll die gesamte Information aller Items, die für die Beurteilung des Schweregrades wesentlich ist, erschöpfend darstellen. Soweit die Information, die in den Ausprägungen der Items repräsentiert wird, nicht im Summenscore abgebildet wird, soll sie irrelevant für die Beurteilung des Schweregrades der Depression bzw. für die relative Ausprägung des Summenscores sein.

Unter den beiden genannten Meßmodellen ist das Modell von RASCH das eleganteste: in ihm sind keine Willkürannahmen enthalten; die Bedingungen "Suffizienz - Summierbarkeit" und



"Stichprobenunabhängigkeit - Transferability" stellen zusammen mit der Bedingung "Homogenität" jeweils hinreichende und notwendige Bedingungen für das Meßmodell dar (17). Das Modell von RASCH ist aber auch restriktiver als das von MOKKEN (30).

Für die weiteren praktischen Anwendungen dieser Modelle ist relevant, daß die entsprechenden Grundmodelle dichotome Itemausprägungen voraussetzen; so sind bei einer Applikation dieser Modelle auf die HAMD Dichotomisierungen der Items notwendig.\*

Zur Prüfung der Verwendbarkeit der HAMD als eine homogene Schweregradsskala wurde

1. die Homogenität im Rahmen der Modells von MOKKEN in allen drei Stichproben ermittelt;
2. ein Modelltest für die Adaptation an das Modell von RASCH durchgeführt (17).

### 3.1 Methoden:

Wir verwendeten die Version mit 17 Items der HAMD, da die Itemanalysen des letzten Abschnitts keine Vorteile der HAMD mit 21 Items gegenüber der kürzeren Version zeigen.

\*Zwar liegen auch RASCH-Modelle für mehrstufige Itemausprägungen vor, doch dies führt zu einer vermehrten Anzahl von Modellparametern, die bei jeweiligen Anwendungen zusätzlich zu schätzen sind. Da die Parameter simultan geschätzt werden, führt dies zu einer Erhöhung des notwendigen Stichprobenumfangs. Zudem müssen für Modellfunktionen in diesem verallgemeinerten Fall relativ willkürliche Annahmen getroffen werden, die jeweils schwer zu rechtfertigen sind. Daher beschränken wir uns auf die Grundmodelle und arbeiten mit Dichotomisierungen der Itemausprägungen.



Die notwendigen Dichotomisierungen der Itemausprägungen wurden in verschiedener Weise durchgeführt, um methodische Artefakte zu vermeiden:

Eine Dichotomisierungsmethode war apriorisch; bei fünfstufigen Itemausprägungen wurden die Ausprägungen 0, 1, 2 als die eine und die Ausprägungen 3, 4 als die andere Antwortkategorie gewertet.

Die andere Dichotomisierungsmethode war aposteriorisch; die Items wurden von ihrem jeweiligen, in den einzelnen Stichproben ermittelten Mittelwerten dichotomisiert; dieses Verfahren wurde von COHEN (12) empfohlen, um möglichst viel Varianz zu konservieren.

Für die Prüfung auf Verträglichkeit mit dem RASCH-Modell sind Kriterien anzugeben, nach denen die untersuchte Stichprobe in zwei Untergruppen zu teilen ist; für jede Untergruppe werden Itemparameter getrennt geschätzt, die dann auf signifikante Unterschiede geprüft werden (16); als Teilungskriterien wurden verwendet:

Zufall,  
Alter (Median),  
Geschlecht,  
Summenscore der HAMD (17 Items) (Median).

### 3.2 Ergebnisse:

1. In allen 3 untersuchten Stichproben erweist sich die HAMD (17 Items) als inhomogen gemessen am LOEVINGER-Koeffizienten (Tabelle 7).

Die Normwerte für die Beurteilung des LOEVINGER-Homogenitätskoeffizienten finden sich in (35); Koeffizienten unter 0,30 gelten als unzureichend (35).

Diese Resultate sind unabhängig von der verwendeten Dichotomisierungsmethode.



Tabelle 7

Homogenitätsanalyse der HAMD  
(Chi-Quadrat-Werte und empirische p-Werte)

signifikante p-Werte bezeichnen Modellabweichungen

| Stichproben                                       |                                         |                                         |                                         |
|---------------------------------------------------|-----------------------------------------|-----------------------------------------|-----------------------------------------|
| Homogenitäts-<br>kriterien                        | P1                                      | P2                                      | P3                                      |
| LOEVINGER-Koeffizient                             | .11 (.09)                               | .12 (.09)                               | .16 (.14)                               |
| Rasch-Modell-Anpassungs-<br>Tests mit Kriterium:* |                                         |                                         |                                         |
| Geschlecht                                        | $\chi^2 = 39,3$ (50,7)<br>p = .00 (.00) | $\chi^2 = 40,9$ (51,3)<br>p = .00 (.00) | $\chi^2 = 45,9$ (48,3)<br>p = .00 (.00) |
| Alter                                             | $\chi^2 = 45,2$ (60,3)<br>p = .00 (.00) | $\chi^2 = 48,1$ (63,0)<br>p = .00 (.00) | $\chi^2 = 40,2$ (56,3)<br>p = .00 (.00) |
| Median der Skala                                  | $\chi^2 = 40,8$ (65,7)<br>p = .00 (.00) | $\chi^2 = 42,6$ (59,3)<br>p = .00 (.00) | $\chi^2 = 35,7$ (60,2)<br>p = .00 (.00) |
| Zufall                                            | $\chi^2 = 20,0$ (21,0)<br>p = .22 (.26) | $\chi^2 = 20,9$ (21,9)<br>p = .26 (.38) | $\chi^2 = 18,3$ (20,5)<br>p = .19 (.24) |

Dichotomisierung: a priori

(Werte in Klammern beziehen sich auf die Version mit 21 Items, die ungeklammerten Werte auf die Version mit 17 Items.)

2. Die Prüfung auf Verträglichkeit mit dem RASCH-Modell zeigt nur bei Zufallsteilung eine ausreichende Modellverträglichkeit (Tabelle 7). Dieses Resultat gilt für alle untersuchten Stichproben und für jede der beiden verwendeten Dichotomisierungsmethoden.

\*Die Testgröße ist Chi-Quadrat-verteilt. Modelltest von ANDERSEN (16).



### 3.3 Diskussion:

Die Homogenitätsanalyse weist in allen drei untersuchten Stichproben unabhängig von der Methode der Dichotomisierung auf die Inhomogenität der HAMD hin. Das gilt für die Homogenitätsanalyse nach MOKKEN und für die Homogenitätsanalyse nach RASCH. Bei der Adaptation an das RASCH-Modell zeigt sich lediglich bei Zufallsteilung der Stichproben eine hinreichende Modellverträglichkeit; dies ist auf die Vergleichbarkeit der bei Zufallsteilung resultierenden Untergruppen zurückzuführen, wenn nur die Stichprobenumfang groß genug ist. Das Verhältnis des Stichprobenumfangs zur Anzahl der Items liegt in jeder Stichprobe höher als 5. Für die Prüfung der Modellverträglichkeit mit dem RASCH-Modell ist dieses Verhältnis in unserem Datensatz ausreichend; ein ungenügender Stichprobenumfang würde wegen mangelnden signifikanten Ergebnissen zu artifiziellen Verträglichkeiten mit dem RASCH-Modell führen (30). Die bei allen nicht-randomisierten Unterteilungen der Stichprobe auftretenden signifikanten Modellabweichungen sprechen daher für einen ausreichenden Stichprobenumfang.

Die Homogenitätsanalysen bekräftigen also den Zweifel an der Verwendbarkeit der HAMD als homogener Schweregradskala, der bei den faktorenanalytischen Untersuchungen aufgetreten ist. Daher ist es notwendig nach Subskalen in der HAMD zu suchen, die bessere Schweregradindikatoren als der Globalsummenscore der HAMD darstellen.



#### 4. REDUKTION DER HAMILTON-DEPRESSIONS-SKALA

Durch die Reduktion der HAMD sollte eine homogene Subskala extrahiert werden. Eine Skala ist dabei homogen, wenn sie die beiden folgenden Homogenitätsbedingungen erfüllt:

1. Jedes Item in der Subskala soll den Schweregrad des depressiven Syndroms beurteilen. Daher ist ein monotoner Zusammenhang zwischen den Itemausprägungen und dem Schweregrad des Syndroms zu fordern. Damit wird verhindert, daß die Intensivierung eines Symptomes, das in einem Item repräsentiert wird, ceteris paribus zu keiner Erhöhung des Globalsummenscores führt; das ist in der HAMD z.B. beim Item Nr. 18 (Tagesschwankungen) der Fall, wenn sich die Intensivierung der depressiven Symptomatik in einer Aufhebung der Tagesschwankung niederschlägt (z.B.(50)).
2. Die Häufigkeit der Items, die eine positive Ausprägung aufweisen, sollten mit dem Schweregrad der Ausprägungen der einzelnen Items in einem positiven Zusammenhang stehen; Patienten, die in einigen Items relativ hohe Ausprägungen aufweisen, sollten auch in den anderen Items relativ hohe Ausprägungen haben. Damit wird verhindert, daß Patienten mit vielen, geringgradig ausgeprägten Symptomen, die nur locker mit dem depressiven Syndrom verknüpft sind (Hypochondrie, fehlende Krankheitseinsicht, Anakasmus, Depersonalisation, somatische Angst), denselben Wert der Globalsumme in der Depressionsskala erhalten wie Patienten, die ausschließlich Kernsymptome (Hemmung, Schuld, Suizidalität) zeigen.

Diese Bedingung kann auch anders formuliert werden: Unterscheiden sich die Ausprägungen eines Items in einer beliebigen Untergruppe einer Stichprobe von den Ausprägungen dieses Items in der Restgruppe dieser Stichprobe, sollen die anderen Items ähnliche Unterschiede aufweisen (Bedingung der Transferability d.h. Stichprobenunabhängigkeit).

Beide Bedingungen sind im RASCH-Modell (16) operationalisiert.



In der Skalierungstheorie von MOKKEN bzw. im Homogenitätskoeffizienten von LOEVINGER ist die erste Bedingung in vollem, und die zweite Bedingung in beschränktem Umfang operationalisiert. Entsprechend ist das RASCH-Modell restriktiver (30). Eine Reduktion der HAMD auf Subskalen, die den Bedingungen des RASCH-Modells genügen, führt daher zu kürzeren Subskalen als die Verwendung der Skalierungstheorie von MOKKEN.

#### 4.1 Methodologische Vorüberlegungen:

##### 1. Wahl der Patienten-Stichprobe:

Die Stichprobe, deren HAMD-Daten zur Reduktion auf eine Subskala dienen sollen, sollte möglichst homogen sein. Denn eine homogene Population läßt eine Skala stets homogener erscheinen als eine heterogene Population (vgl. die in Abschnitt 3 dargestellten Daten); eine Reduktion in einer homogeneren Gruppe führt also zu einer längeren Subskala als in einer heterogenen Gruppe. Um die einzelnen Schritte der Reduktion besser kontrollieren zu können, wird eine homogene Stichprobe gewählt.

Die bei der Reduktion extrahierte Subskala sollte eine zeitunabhängige (d.i. vom Rating-Zeitpunkt unabhängig) Homogenität aufweisen. Um eine möglichst zeitunabhängige Subskala zu erhalten, wird das Reduktionsverfahren auf die Daten einer Stichprobe, die zu zwei verschiedenen Zeitpunkten bzw. unter zwei verschiedenen Bedingungen beurteilt wurde, für jeden Rating-Zeitpunkt getrennt angewendet.

So wurde die Reduktion der HAMD aufgrund der Itemausprägungen in der Stichprobe P1 (ausschließlich Patienten mit endogener Depression RDC vor Aufnahme der Therapie) und in der Stichprobe P3 (ausschließlich Patienten, die auch in der Stichprobe P1 repräsentiert sind, nach Therapie) durchgeführt.

##### 2. Der Reduktion wird die HAMD in der Version mit 17 Items



zugrundegelegt. In Abschnitt 2 wurde nämlich dargelegt, daß die Version mit 21 Items keinen Vorteil gegenüber der Version mit 17 Items aufweist.

3. Die Reduktion wird mit der Skalierungstheorie von MOKKEN durchgeführt. Dieses Verfahren ist weniger restriktiv als Verfahren, die auf der Grundlage des RASCH-Modells arbeiten (30). Daher kann mit dem hier gewählten Verfahren der Prozess der Reduktion besser kontrolliert werden. Wir schließen uns damit einer in der Literatur gegebenen Empfehlung an (30). Die Reduktion erfolgt iterativ, indem aus der Skala bzw. der Subskala jeweils jenes Item eliminiert wird, das die geringste Homogenität aufweist. Das iterative Verfahren stoppt dann, wenn erstmals eine Subskala resultiert, deren sämtliche Items einen LOEVINGER-Homogenitätskoeffizienten größer als 0,30 aufweisen. Homogenitätskoeffizienten größer als 0,30 sind gemäß der Theorie von MOKKEN jedenfalls nicht inhomogen. Skalen mit Homogenitätskoeffizienten größer als 0,30 werden als homogen betrachtet (35).
4. Die Anwendung der Skalierungstheorie von MOKKEN, die die Grundlagen der Reduktion bilden soll, setzt eine dichotomisierte Itemstruktur voraus (35). Eine Dichotomisierung der HAMD-Items, die 5, 4 oder 3 Ausprägungsgrade aufweisen, kann in verschiedener Weise geschehen:  
durch eine Apriori-Dichotomisierung zwischen den Ausprägungsstufen 1 und 2 (resp. 2 und 3) bei allen Items mit 4 oder 3 Ausprägungsstufen (resp. 5 Ausprägungsstufen) oder durch Aposteriori-Dichotomisierung am Mittelwert oder Median der empirisch erhobenen Itemausprägungen.

Um zwischen Items mit möglichst unterschiedlichen mittleren Ausprägungen unterscheiden zu können, empfiehlt sich eine Apriori-Dichotomisierung, wie sie eben beschrieben wurde; dies ist nämlich notwendig, um Mittelwertunterschiede zwischen Items zu erhalten.



#### 4.2 Ergebnisse:

##### 1. Stichprobe P1 (Tabelle 8)

Die resultierende Subskala (LOEVINGER-Homogenitätskoeffizient 0,31) beinhaltet sechs Items:

Item Nr. 1 (depressive Stimmung)

Item Nr. 2 (Schuldgefühle)

Item Nr. 7 (Arbeit und sonstige Tätigkeiten)

Item Nr. 8 (Hemmung)

Item Nr. 9 (Erregung)

Item Nr.10 (Angst - psychisch).

Das Item Nr.13 (körperliche Symptome - allgemein) wird dabei als das letzte im iterativen Verfahren ausgeschlossen; würde die extrahierte Subskala um dieses Item ergänzt werden, sänke der Homogenitätskoeffizient auf 0.27.

##### 2. Stichprobe P3 (Tabelle 9)

Als Subskala (Homogenitätskoeffizient 0,35) wurden sieben Items extrahiert:

Item Nr. 1 (depressive Stimmung)

Item Nr. 2 (Schuldgefühle)

Item Nr. 7 (Arbeit und sonstige Tätigkeit)

Item Nr. 8 (Hemmung)

Item Nr. 9 (Erregung)

Item Nr.10 (Angst - psychisch)

Item Nr.13 (Körperliche Symptome - allgemein).

Das Item Nr. 12 (Angst - somatisch) wurde als letztes Item eliminiert, bevor die sieben Items umfassende Subskala resultierte. Würde dieses Item ergänzt werden, so sänke der Homogenitätskoeffizient auf 0,27.

Die in der Stichprobe P1 extrahierte Subskala mit sechs Items hat in der Stichprobe P3 einen hohen Homogenitätskoeffizienten von 0,38.



#### 4.3 Diskussion:

In der Stichprobe P1 wurde eine kürzere Skala extrahiert als in der Stichprobe P3; die in P1 ermittelte Subskala umfaßt ausschließlich Items, die auch in der Subskala enthalten sind, die in der Stichprobe P3 ermittelt wurden. In der Stichprobe P3 erscheinen Skalen homogener als in der Stichprobe P1. Dies kann auf den geringeren Stichprobenumfang in P3 oder auf eine Verbesserung der Modellverträglichkeit im Therapieverlauf zurückgeführt werden.

Die Subskala, die zu verschiedenen Rating-Zeitpunkten hinreichend homogen ist, ist also die in der Stichprobe P1 ermittelte Subskala mit den sechs Items: "depressive Stimmung", "Schuldgefühle", "Arbeit und sonstige Tätigkeiten", "Hemmung", "Erregung", "psychische Angst". Die Subskala besteht also aus Items, die

- a) den Kernbereich depressiver Symptomatik repräsentieren,
- b) keine somatischen Beschwerden abbilden.

Unter den in der Literatur als typisch für die endogene Depression angeführten Symptomen fehlen jedoch die Vitalstörungen (z.B. (46)). Diese sind nur in einem Item der HAMD (Item Nr. 13 "Körperliche SYmptome - allgemein") zusammen mit anderen Symptomen, die vorwiegend somatische Beschwerden erfassen, repräsentiert. Entsprechend ist nicht festzustellen, ob allgemeine körperliche Beschwerden oder Vitalstörungen für die relative Inhomogenität des entsprechenden Items verantwortlich sind. Für eine Stichprobe mit Patienten mit larvierter Depression (27) würde dieses Resultat möglicherweise nicht resultieren; die Zielpopulation dieser Untersuchung ist jedoch die Gesamtheit der stationär behandelten depressiven Patienten, die an Evaluationsstudien teilnehmen.

Dieses Ergebnis unterstützt die These, daß die Bewertung von Symptomen aus dem depressiven Kernbereich den Schweregrad



depressiver Syndrome besser darstellen kann als die Bewertung von Symptomen, die nur lose mit dem depressiven Syndrom assoziiert sind oder körperliche Symptomatik repräsentieren (22). Damit wird verhindert, daß ein Patient mit vielen Symptomen, die nur lose mit dem depressiven Syndrom assoziiert sind und mit wenig Symptomen aus dem Kernbereich der depressiven Symptomatik dieselbe Ausprägung erhält, wie ein Patient, der ausschließlich Symptome aus dem Kernbereich der depressiven Symptomatik in relativ hoher Ausprägung aufweist. Das Fehlen von Items, die körperliche Beschwerden beschreiben, zeigt an, daß in den untersuchten Stichproben die Ausprägungen körperlicher Mißempfindungen durch andere Faktoren beeinflußt werden als die Ausprägungen von Symptomen aus dem Kernbereich depressiver Symptomatik.

Das Fehlen von Items in der extrahierten Subskala, die die körperliche Symptomatik repräsentieren, bewirkt, daß die Symptome der körperlichen Nebenwirkungen der antidepressiven Therapie nicht in die Beurteilung des Schweregrads des depressiven Syndroms eingehen. Von MONTGOMERY und ASBERG (36) wurde gegen die 17-Items-Version der HAMD vorgebracht, daß die Beurteilung des Schweregrades der depressiven Symptomatik mit der Beurteilung von Nebenwirkungen unter antidepressiver Therapie untrennbar verbunden ist; dieses Argument gilt für die extrahierte Subskala der HAMD nicht mehr.

BECH et al. (4) hat in einer nosologisch heterogenen Gruppe depressiver Patienten mit einer Methode (Varianzanalyse, Verwendung der Globalbeurteilung der depressiven Symptomatik) eine Subskala extrahiert, die ebenso 6 Items umfaßt. Die Subskala stimmt in 5 Items mit der hier extrahierten Subskala überein; diese Ähnlichkeit ist überraschend. In der Analyse von BECH et al. (2) wurde das Item Nr. 9 (Erregung), das in der von uns extrahierten Subskala vorkommt, wegen mangelnder Varianz nicht in die Analyse einbezogen. In dem hier untersuchten Datensatz zeigt dieses Item zwar auch eine relativ



geringe Standardabweichung; für die Homogenitätsanalyse ist aber kein unterer Schwellenwert für die Standardabweichung von Items Voraussetzung. In der Subskala von BECH et al. wurde an Stelle von Item Nr. 9 das Item Nr. 13 (körperliche Symptome - allgemein) aufgenommen. Dieses Item ist das letzte, das bei iterativer Reduktion der HAMD in der Stichprobe P1 eliminiert wurde, bevor die homogene Subskala mit sechs Items resultierte; in der Stichprobe P3 ist dieses Item in der resultierenden homogenen Subskala enthalten. Item Nr. 13 (körperliche Symptome - allgemein) ist also relativ - wenngleich nicht hinreichend - homogen bezüglich der hier isolierten homogenen Subskala. Diese Diskrepanz zwischen der hier isolierten Subskala und der von BECH et al. (6) ist auf Unterschiede in der Stichprobenzusammensetzung, in der verwendeten Methode der Skalenreduktion oder auf Zufallsschwankungen zurückzuführen.

Bei der Isolierung von Subskalen können Zufallsfehler vorkommen; das gilt für die Skala von BECH et al. und die hier isolierte Subskala. Im Reduktionsverfahren von MOKKEN, das in dieser Arbeit verwendet wurde, steigt zudem mit jedem Prüfungsschritt der Einfluß der Zufallsfehler (capitalizing of chance (35), Seite 196). Da der Unterschied zwischen der hier und der von BECH et al. isolierten Subskala lediglich in dem Item besteht, das beim letzten Konstruktionsschritt extrahiert wurde, ist der Einfluß des Zufallsfehlers als gering anzusehen.



## 5. WEITERE REDUKTION DER HAMILTON-DEPRESSIONS-SKALA

Ausgangspunkt für eine weitere Reduktion sind die folgenden Überlegungen:

1. Die reduzierte Subskala, die hier isoliert wurde, und die von BECH in einer anders zusammengesetzten Stichprobe mit anderen Mitteln (Varianzanalyse anstelle der Homogenitätsanalyse nach MOKKEN) extrahiert wurde, decken sich in fünf Items und unterscheiden sich nur jeweils in einem Item; daher liegt es nahe, zu prüfen, ob die fünf Items umfassende Subskala - die in beiden genannten Reduktionen der HAMD - enthalten ist, den hier verwendeten Optimierungskriterien besser entspricht.
2. In der Subskala von BECH ist Item Nr. 13 (körperliche Symptome - allgemein) das (gemäß dem RASCH-Modell) inhomogenste Item; in einigen Analysen genügt dieses Item den Bedingungen des Modells nicht. So liegt es nahe, jene Skala zu analysieren, die sich aus der Subskala von BECH durch Eliminierung von Item Nr. 13 ergibt.
3. In der hier durchgeführten Reduktion der HAMD ist Item Nr. 9 (Agitation) das inhomogenste Item, denn:
  - a) es weist die geringste Verträglichkeit mit dem RASCH-Modell bei Validierungsuntersuchungen auf (siehe Kapitel 4);
  - b) Bei der Isolierung der Subskala in Kapitel 3 weist dieses Item die geringste Homogenität (gemäß dem LOEVINGER-Koeffizienten) unter den Items der extrahierten Subskala auf (Kapitel 3); damit wird eine Überprüfung jener Subskala nahegelegt, die sich aus der in Kapitel 3 extrahierten durch Weglassen von Item Nr. 13 ergibt.

Alle drei Überlegungen führen zu derselben Subskala , welche fünf Items umfaßt:

Item Nr. 1 (depressive Stimmung),  
Item Nr. 2 (Schuldgefühle),



Item Nr. 7 (Arbeit und sonstige Tätigkeiten),  
Item Nr. 8 (Hemmung),  
Item Nr.10 (psychische Angst).

In den Stichproben P1 und P3, aus denen die Subskala in Kapitel 4 gewonnen wurde, werden die drei Subskalen (eine gemäß Kapitel 4, die von BECH et al. vorgeschlagene und der Überlappungsbereich dieser beiden Subskalen von Itemumfang 5) bezüglich ihrer relativen Nutzen verglichen. Ziel ist, aus den drei Subskalen eine auszuwählen, die den Kriterien der Homogenität und der Populationsunabhängigkeit am besten genügt.

#### 5.1 Methoden:

Die drei Subskalen werden bezüglich der erwünschten Eigenschaften Homogenität (LOEVINGER-Koeffizient, Schweregradfaktor in der Faktorenanalyse und Anpassungstests für das RASCH-Modell) und der Populationsunabhängigkeit (Anpassungstests für das RASCH-Modell) untersucht. Der Zusammenhang der Subskalen mit der 17-Items-Version der HAMD wird ebenfalls angegeben. Um zeitliche Stabilität für die getroffenen Aussagen zu erreichen, werden die entsprechenden Kennwerte für die Stichproben P1 und P3 erfaßt (P3 umfaßt Patienten, die auch in P1 enthalten sind, die jedoch 3 Wochen nach Therapiebeginn beurteilt werden).



## 5.2 Ergebnisse:

Tabelle 8

Statistische Kennwerte und Homogenitätsanalysen  
Stichprobe: P1 (n = 108)

|                                                            | Subskala 1<br>(gem. Kap. 4) | Subskala 2<br>(gem. BECH et al.) | Subskala 3<br>(Itemumfang 5) |
|------------------------------------------------------------|-----------------------------|----------------------------------|------------------------------|
| Mittelwert                                                 | 10,7                        | 11,9                             | 9,6                          |
| Varianz                                                    | 3,9                         | 4,6                              | 3,5                          |
| Innere Konsistenz<br>(Alpha-Koeffizient<br>CRONBACH)       | .69                         | .69                              | .65                          |
| LOEVINGER-Koeffizient                                      | .31                         | .28                              | .35                          |
| RASCH-Modell-<br>Anpassungstests*<br>nach Kriterien:       |                             |                                  |                              |
| Alter                                                      | +                           | +                                | +                            |
| Geschlecht                                                 | +                           | +                                | +                            |
| Median der Skala                                           | -                           | -                                | (+)                          |
| Zufall                                                     | ++                          | ++                               | ++                           |
| Schweregradfaktor<br>(Generalfaktor in<br>Faktorenanalyse) | +                           | -                                | +                            |
| SPEARMAN-Korrel.<br>mit HAMD (17 Items)                    | .69                         | .71                              | .73                          |
| KENDALL-Korrel.<br>mit HAMD (17 Items)                     | .71                         | .71                              | .74                          |

Subskala 1: Items Nr. 1, 2, 7, 8, 9, 10 (Itemumfang 6)

Subskala 2: Items Nr. 1, 2, 7, 8, 10, 13 (Itemumfang 6)

Subskala 3: Items Nr. 1, 2, 7, 8, 10 (Itemumfang 5)

\* +: signifikante Modellabweichung ( $0,01 < p \leq 0,05$ )

++: signifikante Modellabweichung ( $0,01 \leq p$ )



Tabelle 9

Statistische Kennwerte und Homogenitätsanalysen  
Stichprobe: P3 (n = 88)

|                                                            | Subskala 1<br>(gem. Kap. 4) | Subskala 2<br>(gem. BECH et al.) | Subskala 3<br>(Itemumfang 5) |
|------------------------------------------------------------|-----------------------------|----------------------------------|------------------------------|
| Mittelwert                                                 | 4,4                         | 5,4                              | 3,7                          |
| Varianz                                                    | 4,0                         | 4,3                              | 3,9                          |
| Innere Konsistenz<br>(Alpha-Koeffizient<br>CRONBACH)       | .72                         | .73                              | .70                          |
| LOEVINGER-Koeffizient                                      | .28                         | .29                              | .32                          |
| RASCH-Modell-<br>Anpassungstests*<br>nach Kriterien:       |                             |                                  |                              |
| Alter                                                      | +                           | +                                | +                            |
| Geschlecht                                                 | +                           | +                                | +                            |
| Median der Skala                                           | (+)                         | (+)                              | +                            |
| Zufall                                                     | ++                          | ++                               | ++                           |
| Schweregradfaktor<br>(Generalfaktor in<br>Faktorenanalyse) | +                           | (+)                              | +                            |
| SPEARMAN-Korrel.<br>mit HAMD (17 Items)                    | .69                         | .73                              | .75                          |
| KENDALL-Korrel.<br>mit HAMD (17 Items)                     | .69                         | .73                              | .72                          |

Subskala 1: Items Nr. 1, 2, 7, 8, 9, 10 (Itemumfang 6)

Subskala 2: Items Nr. 1, 2, 7, 8, 10, 13 (Itemumfang 6)

Subskala 3: Items Nr. 1, 2, 7, 8, 10 (Itemumfang 5)

\* +: signifikante Modellabweichung ( $0,01 \leq p \leq 0,05$ )

++: signifikante Modellabweichung ( $0,01 \leq p$ )



Die untersuchte Subskala mit 5 Items zeigt eine hohe Korrelation (siehe Tabelle 8,9) sowohl mit den anderen beiden Subskalen als auch mit der gesamten HAMILTON-Skala (17 Items). Insgesamt ist der korrelative Zusammenhang der Skala mit fünf Items und der 17-Items-Version der HAMD vergleichbar mit dem korrelativen Zusammenhang zwischen den Subskalen mit Itemumfang sechs und der 17-Items-Version der HAMD.

Für die Koeffizienten der inneren Konsistenz nach CRONBACH ergeben sich für die Subskala vom Itemumfang fünf vergleichbare, aber leicht niedrigere Werte, als für die Subskalen mit Itemumfang sechs; die innere Konstistenz der Subskala vom Itemumfang fünf ist jedoch nicht geringer als für die 17-Items-Version der HAMD. Da der Koeffizient von CRONBACH auch als Reliabilitätskoeffizient interpretiert werden kann (42), ergibt sich der überraschende Befund, daß die HAMD in ihrer vollen Version (17 Items bzw. 21 Items) weniger reliabel (im Sinne der Test-Retest-Reliabilität) als die drei Subskalen der HAMD.

Die drei Subskalen zeigen weiterhin eine deutlich höhere Homogenität und eine deutlich höhere Populationsunabhängigkeit als die 17-Items-Version der HAMD. Für die Subskala vom Itemumfang fünf Items ergibt sich weiterhin eine Überlegenheit (bezüglich Homogenität und Stichprobenunabhängigkeit) im Vergleich zu den beiden Subskalen vom Itemumfang sechs Items.

Somit ist die Subskala mit fünf Items zur Beurteilung des Schweregrades der depressiven Symptomatik besser geeignet, als alle vorher diskutierten Versionen der HAMD. Gleichwohl sind die Ergebnisse nicht in vollem Umfange befriedigend, denn bei der Prüfung auf die Verträglichkeit mit dem RASCH-Modell zeigen sich auf Item-Niveau bei Teilung der Stichprobe am Median des Summenwertes für die Fünf-Items-Subskala signifikante Modellabweichungen; diese betreffen ausschließlich Item Nr. 2.

Dieses Resultat gilt für die beiden Stichpunkte P1 und P3



gleichermaßen. Allerdings sind diese Modellabweichungen in keinem Fall so ausgeprägt, daß sie auf Skalenniveau zu einer signifikanten Abweichung von den Bedingungen des RASCH-Modells führen würden.

Bei der Überprüfung der Homogenität mit der Faktorenanalyse (Hauptkomponentenanalyse) ergibt sich für die Version mit 5 Items und für die in Kapitel 4 isolierte Subskala mit 6 Items ein hinreichendes Ergebnis: Es resultiert eine Ein-Faktoren-Lösung nach dem Scree-Graph-Kriterium von CATTELL (8), wobei der extrahierte Faktor ein Generalfaktor (erklärte Varianz 65 %) ist, der auf allen Items relevant lädt (Ladungskoeffizienten größer 0,40). Dieses Ergebnis gilt für die beiden Stichproben P1 und P3. Für die von BECH et al. isolierte Subskala mit sechs Items ergibt sich zwar auch eine Ein-Faktoren-Lösung, der extrahierte Faktor ist jedoch weder in der Stichprobe P1 noch in der Stichprobe P3 ein Generalfaktor; denn auf dem Item Nr. 13 "körperliche Symptome - allgemein" lädt dieser extrahierte Faktor stets niedriger als 0,35.

Die Überlegenheit der Subskala mit fünf Items über die beiden Subskalen mit sechs Items ist in den Stichproben P1 und P3 deutlich. Der Vorzug, der der Skala mit Itemumfang fünf gegeben werden muß, ist also frei von Schwankungen, die durch den Therapieverlauf induziert sind.

### 5.3 Diskussion:

Die Subskala mit fünf Items ist in allen untersuchten Stichproben eine umfassend homogene Skala; diese wird durch Faktorenanalyse, RASCH-Analysen und LOEVINGER-Koeffizienten, die sich für diese Skala ergeben, bestätigt. Zudem erweist sich diese Skala als populationsunabhängig; ihre Populationsunabhängigkeit ist besser als die der beiden anderen Subskalen und die der HAMD (17 Items, 21 Items).

Der Koeffizient für die innere Konsistenz (CRONBACH) ist in der Fünf-Item-Subskala allerdings nicht höher als in den



Versionen mit sechs Items; dies ist auf die Reduktion des Itemumfangs zurückzuführen.

Die Ergebnisse der RASCH-Analyse ergeben zwar für alle untersuchten Teilungskriterien der Stichproben P1 und P3 für die Version mit fünf Items bessere Ergebnisse als für die Versionen mit sechs Items. Gleichwohl sind die Ergebnisse für die Fünf-Items-Subskala nicht in vollem Umfang befriedigend, denn auf dem Item-Niveau kommt es bei Teilung der Stichprobe am Median des Summenwertes zu Modellabweichungen.

Diese betreffen ausschließlich Item Nr. 2 (Schuldgefühle). Sie treten in beiden Stichproben P1 und P3 auf. Allerdings sind diese Modellabweichungen in keinem Fall so ausgeprägt, daß sie auf Skalenniveau zu einer signifikanten Abweichung von den Bedingungen des RASCH-Modells führen würden. Entsprechend gehen wir davon aus, daß die Fünf-Item-Subskala den Bedingungen des RASCH-Modells genügt.

Eine weitere Reduktion der Skala mit fünf Items empfiehlt sich nicht, da sich bei der Reduktion der Version mit sechs Items auf die Subskala mit fünf Items für den Koeffizienten der inneren Konsistenz Wertminderungen ergeben haben; wir interpretieren diese Wertminderung als eine Reduktion der Reliabilität (Test, Retest). Diese sind jedoch nicht so nachhaltig, daß die Reliabilität (gemessen am Alpha-Koeffizienten) der Subskala mit fünf Items deutlich niedriger als in den Versionen mit sechs Items wäre. Sie liegt weiterhin höher als in den Versionen mit 17 oder 21 Items. Den partiellen Reliabilitätsverlust (gegenüber der Version mit sechs Items) für die Fünf-Item-Subskala nehmen wir jedoch als Hinweis, daß eine weitere Reduktion der HAMD nur unter weiteren Reliabilitätsverlusten (gemessen am Alpha-Koeffizienten) erreicht werden kann. Daher unterlassen wir eine weitere Reduktion der HAMD.

Aufgrund der Untersuchungen in den Stichproben P1 und P3 ist die Subskala mit fünf Items



- a) eine bessere Skala zur Beurteilung des Schweregrades der depressiven Symptomatik als die HAMD (17 oder 21 Items) (gemessen an den hier zugrunde gelegten Prüfkriterien),
- b) unter den aus der HAMD isolierten Subskalen diejenige, die den genannten Verwendungszwecken (Homogenität und Populationsunabhängigkeit) am ehesten entspricht.

Ungeprüft mußte bleiben, ob die Reliabilität (im Sinne der Interrater-Reliabilität) der Subskala mit fünf Items ausreichend ist bzw. der Reliabilität der HAMD mit 17 bzw. 21 Items entspricht.

Im folgenden Kapitel wird eine Validierung der Subskalen der HAMD aufgrund der Daten der Stichprobe P2 durchgeführt.



## 6. VALIDIERUNG DER EXTRAHIERTEN SUBSKALEN

Zur Validierung ist die Analyse der Daten der reduzierten Subskalen (bestehend aus Items Nr 1,2,7,8,9,10 bzw. Items Nr. 1,2,7,8,10) der HAMD in anderen als der zur Subskalenkonstruktion verwendeten Patientengruppen nötig. Die Validierung sollte den Zusammenhang mit anderen Instrumenten zur Beurteilung des Schweregrades depressiver Symptomatik und die Homogenität der Subskala nach den folgenden Gesichtspunkten prüfen:

1. Es ist zu prüfen, ob alle Items gleichermaßen als Maßstäbe des Schweregrades depressiver Symptomatik gelten können.
2. Es ist zu prüfen, ob die Skalen populationsunabhängig stets dieselbe Qualität beurteilen; insbesondere ist zu prüfen ob
  - a) der Schweregrad der depressiven Symptomatik,
  - b) patientenbezogene Daten (Geschlecht, Alter)die Ausprägungen der verschiedenen Items in unterschiedlicher Weise beeinflussen.

Die beiden Forderungen 1. und 2. bilden die Grundlage der probabilistischen Meßmodelle von RASCH und MOKKEN.

Die regelmäßig in der Literatur erhobene Forderung nach einer möglichst hohen Reliabilität (in Form von Interrater-Reliabilität) wird hier nicht angeführt, da sie aufgrund des settings nicht hinreichend geprüft werden konnte (keine zwei unabhängigen Beurteiler). Lediglich die Prüfung der inneren Konsistenz konnte vorgenommen werden; diese kann als Reliabilität (in Form von Test-Retest-Reliabilität) interpretiert werden (29).

Zur Prüfung der genannten Validierungsbedingungen werden die folgenden Kennwerte dargestellt:

1. Item-Trennschärfe,
2. innere Konsistenz (Alpha-Koeffizient nach CRONBACH),
3. korrelativer Zusammenhang mit der Globalbeurteilung, des Schweregrades,



4. Darstellung des Zusammenhangs zwischen Globalbeurteilung und Globalsummenskore der Subskalen,
5. Homogenitätskoeffizient nach LOEVINGER,
6. Anpassungstest für das RASCH-Modell,
7. Versuch der Darstellung eines Generalfaktors im Rahmen der Faktorenanalyse bzw. Hauptkomponentenanalyse.

Bei der Validierung einer Skala bzw. Subskala muß auch über den Grad der Abhängigkeit von methodischen Bedingungen bei der Konstruktion bzw. Extraktion der Skalen entschieden werden. Hier gehen als solche Bedingungen insbesondere ein:

Wahl der Prüfstichprobe,

Methode der Dichotomisierung der mehrstufigen Items im Rahmen der Homogenitätsanalysen.

Um eine möglichst methodenunabhängige Subskala zu erhalten, wurde die Validierungsstichprobe nach anderen Selektionskriterien erhoben als die Stichprobe, die bei der Extraktion der Subskala mit 6 Items verwendet wurde; anstelle der homogenen Population P1 wurde die heterogenere Population P2 als Validierungsstichprobe verwendet.

Auch die für die Durchführung der Homogenitätsanalysen notwendigen Dichotomisierungen auf Itemebene wurden bei der Validierung nach einem anderen Prinzip vorgenommen als bei der Extraktion der Subskala. Bei der Validierung wurde die Dichotomisierung für jedes Item am Mittelwert vorgenommen; damit wird möglichst viel Varianz konserviert (12).

Die Validierung einer Subskala erfordert die Darstellung des korrelativen Zusammenhangs mit der Globalbeurteilung des Schweregrades der in der Skala abgebildeten Qualität und mit Konkurrenzskalen, die ebenfalls den Schweregrad derselben Qualität beurteilen. Daher wurden in der Stichprobe P2 gleichzeitig mit der HAMD die folgenden Beurteilungsinstrumente erhoben:



Globalbeurteilung des Schweregrades der depressiven Symptomatik auf einer 7-stufigen Skala,  
Beurteilung des Schweregrades der depressiven Symptomatik nach den validierenden Skalen von BECH und RAFAELESEN (MES)(42) und von MONTGOMERY und ASBERG (MADRS)(36).

### 6.1 Methoden:

Der Validierung wurden die Itemausprägungen in den beiden Subskalen (Items Nr. 1, 2, 7, 8, 9, 10 in einer Subskala und Items Nr. 1, 2, 7, 8, 10 in der anderen Subskala) in der Stichprobe P2 zugrunde gelegt. Diese Patientengruppe und die Erhebung der HAMD-Daten sind bereits oben dargestellt worden.

Zudem wurden zum selben Zeitpunkt die beiden Skalen MES und MADRS von gleichen Beurteiler erhoben; dieser beurteilte auch den globalen Schweregrad der Symptomatik.

Die untersuchten Patienten erhielten von ihrem betreuenden Arzt Diagnosen nach dem ICD-System (13); der betreuende Arzt war unabhängig vom Rater der Depressionsskalen. Die ICD-Diagnose wurde verwendet, um die untersuchte Stichprobe zur Prüfung der Abweichung vom RASCH-Modell in zwei Untergruppen zu teilen.

Die beiden Subskalen (Items Nr. 1, 2, 7, 8, 9, 10 in einer Subskala und Items Nr. 1, 2, 7, 8, 10 in der anderen Subskala) wurden folgenden Analysen unterworfen:

Bestimmung der mittleren Trennschärfe in jeder Skala, Ermittlung des Konsistenzkoeffizienten nach CRONBACH in jeder Skala,

Rangkorrelationen nach SPEARMAN und KENDALL

jeder der beiden Subskalen mit der Globalbeurteilung der Symptomatik, der MES und der MADRS,



Ermittlung der mit der Globalbeurteilung nicht positiv korrelierenden Items in jeder Skala,

Bestimmung des LOEVINGER-Homogenitätskoeffizienten in jeder Skala,

Ermittlung von signifikanten Abweichungen vom RASCH-Modell nach dem Modellanpassungstest aufgrund der Teilung der Population nach Zufall (Halbierung), Alter (Teilung der Stichprobe am Median des Alters), Geschlecht,

Diagnose (endogene Depression nach ICD versus andere Diagnose nach ICD);

zusätzlich wurde eine Hauptkomponentenanalyse der beiden Subskalen zur Ermittlung eines Generalfaktors durchgeführt.

Diese Ergebnisse werden verglichen mit den entsprechenden Resultaten für die 17-Items-Version der HAMD. Die genannten Analysen wurden zudem für die Subskala von BECH et al. (Items Nr. 1, 2, 7, 8, 10, 13) durchgeführt und dargestellt.

## 6.2 Ergebnisse (Tabelle 10)

Das Maß für die innere Konsistenz (Alpha-Koeffizient) zeigt in der Validierungsstichprobe für die HAMD (17 Items) und die beiden Subskalen vom Itemumfang sechs (Skala 2, Skala 3) identische Werte; sie genügen den in ((29), Seite 128) angegebenen Bedingungen; der entsprechende Koeffizientenwert für die auf fünf Items reduzierte Skala (Skala 4) zeigt gegenüber den Werten der anderen HAMD-Skalen einen leichten Abfall; trotzdem genügt auch dieser Wert noch den in (29), Seite 128 angegebenen Bedingungen. Die errechneten Werte für die mittlere Trennschärfe zeigen dagegen für die Subskalen (Skala 2, Skala 3, Skala 4) deutlich höhere Werte als für die 17-Items-Version der HAMD. Der Koeffizient für die Homogenität (LOEVINGER) zeigt ebenfalls für die 17-Items-Version der HAMD eine niedrigere Ausprägung als für die Subskalen (Skala 2, Skala 3, Skala 4); während nach den in (35) gegebenen Normwerten die



Tabelle 10

Validierung der HAMD-Subskalen  
Stichprobe: P2 (n = 98)

|                                                                        | HAMD<br>(17 Items) | Subskala 1<br>(gem. Kap. 4) | Subskala 2<br>(gem. BECH et al.) | Subskala 3<br>(Itemumfang 5) |
|------------------------------------------------------------------------|--------------------|-----------------------------|----------------------------------|------------------------------|
| Mittelwert                                                             | 18,19              | 8,9                         | 10,0                             | 7,6                          |
| Varianz                                                                | 7,34               | 4,3                         | 4,9                              | 4,0                          |
| Innere Konsistenz<br>(Alpha-Koeffizient<br>CRONBACH)                   | .72                | .68                         | .69                              | .65                          |
| LOEVINGER-Koeffizient                                                  | .22                | .30                         | .30                              | .38                          |
| RASCH-Modell-<br>Anpassungstests*<br>nach Kriterien:                   |                    |                             |                                  |                              |
| Alter                                                                  | -                  | +                           | +                                | -                            |
| Geschlecht                                                             | -                  | +                           | +                                | -                            |
| Median der Skala                                                       | -                  | +                           | +                                | +                            |
| Schweregrad (global)                                                   | -                  | +                           | +                                | -                            |
| Diagnose                                                               | -                  | -                           | -                                | +                            |
| Zufall                                                                 | ++                 | ++                          | ++                               | -                            |
| Schweregradfaktor<br>(Generalfaktor in<br>Faktorenanalyse)             | -                  | +                           | -                                | +                            |
| Korrelation mit<br>Konkurrenzskalen:<br>(SPEARMANN)<br>HAMD (17 Items) | 1,00               | .82                         | .83                              | .81                          |
| MADRS                                                                  | .85                | .80                         | .79                              | .79                          |
| MES                                                                    | .73                | .83                         | .84                              | .85                          |
| globale Schweregrad-<br>beurteilung                                    | .71                | .75                         | .76                              | .78                          |

Subskala 1: Items Nr. 1, 2, 7, 8, 9, 10 (Itemumfang 6)

Subskala 2: Items Nr. 1, 2, 7, 8, 10, 13 (Itemumfang 6)

Subskala 3: Items Nr. 1, 2, 7, 8, 10 (Itemumfang 5)

\* +: signifikante Modellabweichung ( $0,01 \leq p \leq 0,05$ )

++: signifikante Modellabweichung ( $0,01 \leq p$ )



17-Items-Version der HAMD eine inhomogene Skala ist, sind die Subskalen von geringer, aber ausreichender Homogenität.

Bei der Prüfung mit dem RASCH-Modell ergeben sich folgende Resultate: während die 17-Items-Version der HAMD den Modellbedingungen nur bei Zufallsteilung genügt, zeigen die beiden Subskalen mit Itemumfang sechs (Skala 2, Skala 4) nur bei Teilung der Stichprobe nach den Kriterien "Median der untersuchten Skala" und "Diagnose (endogene Depression versus andere Diagnosen)" signifikante Modellabweichungen; die Subskala mit Itemumfang fünf (Skala 4) ist dagegen bei allen Teilungen der Stichprobe modellkonform; bei der Teilung der Stichprobe nach dem Kriterium "Median der untersuchten Skala" ergibt sich allerdings ein grenzwertiger Befund (empirische Irrtumswahrscheinlichkeit 0,01).

Bei den Faktorenanalysen ergibt sich für die 17-Items-Version der HAMD und für die Subskala von BECH et al. keine Generalfaktorlösung (Hauptkomponentenanalyse und Hauptfaktorenanalyse). Für die Korrelationen mit konkurrierenden Beurteilungsinstrumenten zur Erfassung des Schweregrades der depressiven Symptomatik ergeben sich keine einheitlichen Resultate: während der Zusammenhang mit der Globalbeurteilung und mit der Skala MES für die 17-Items-Version der HAMD weniger stringent ist als mit den Subskalen, ist der Zusammenhang mit der MADRS für die 17-Items-Version der HAMD enger als mit allen Subskalen der HAMD; der Zusammenhang mit konkurrierenden Beurteilungsinstrumenten ist für alle Subskalen der HAMD vergleichbar eng. Die ermittelten Korrelationswerte sind als ausreichend zu betrachten (26).

Auf Item-Niveau ergeben sich für die 17-Items-Version zwei Items mit negativer SPEARMAN-Korrelation mit der Globalbeurteilung der Symptomatik. Für keine der Subskalen sind Items mit solchen negativen Korrelationen feststellbar.



Auf Item-Niveau ergeben sich in Skala 1 (17-Items-Version) für 10 Items Modellunverträglichkeiten bei einer der geprüften Teilungsbedingungen (Signifikanzniveau 5%). In Skala 2 ergeben sich für Item Nr. 9 (Erregung) bei drei und für Item Nr. 2 (Schuldgefühle) bei zwei der geprüften Teilungsbedingungen Modellunverträglichkeiten, die jedoch nur in zwei Fällen zu Modellunverträglichkeiten der gesamten Subskala führen. In Skala 3 ergeben sich für Item Nr. 13 "körperl. Symptome - allgemein" bei 3 und für Item Nr. 2 "Schuldgefühle" bei einer der geprüften Teilungsbedingungen Modellunverträglichkeiten, die jedoch nur in zwei Fällen zu Unverträglichkeiten der gesamten Subskala führen. In Skala 4 ergeben sich für Item Nr. 2 "Schuldgefühle" bei zwei der geprüften Teilungsbedingungen Modellunverträglichkeiten, die jedoch zu keiner Modellabweichung auf dem Skalenniveau führen.

### 6.3 Diskussion:

In allen erhobenen Validierungskriterien ergeben sich für die Subskalen der HAMD (Items Nr. 1,2,7,8,9,10 bzw. Nr. 1,2,7,8,10,13 bzw. Items Nr. 1,2,7,8,10) günstigere Resultate als für die 17-Items-Version der HAMD. Auf der Basis dieser Validierungskriterien sind die Subskalen bessere Instrumente zur Erfassung des Schweregrades der depressiven Symptomatik. Besonders ist hervorzuheben, daß die Subskalen homogener sind, in ihrer inneren Struktur weniger stichprobenabhängig sind und ökonomischer sind. Allerdings konnte die Interrater-Reliabilität nicht erhoben werden; deshalb ist diese Konsequenz nur unter Vorbehalt zu ziehen.

Beim Vergleich der drei Subskalen der HAMD zeigt die kürzeste Version (Itemumfang fünf) bis auf den Alpha-Koeffizienten zur Beurteilung der inneren Konsistenz bessere Resultate als die beiden anderen Subskalen mit Itemumfang sechs. Insbesondere ist die Abhängigkeit der inneren Struktur von der Zusammen-



setzung der Stichprobe für die Skala mit Itemumfang fünf geringer als bei den beiden anderen Subskalen. Daher ist der Skala mit Itemumfang fünf der Vorzug zu geben. Aber auch diese Konsequenz ist nur unter dem Vorbehalt möglich, daß die Interrater-Reliabilität für die Skala mit Umfang fünf nicht ungünstiger als für die Skalen mit Itemumfang sechs ausfällt.

Auf Itemebene zeigt die Analyse der Verträglichkeit mit dem RASCH-Modell, daß bei den Skalen mit Itemumfang sechs die Items Nr. 9 "Erregung" (Skala 2) bzw. Nr. 13 "körperliche Symptome - allgemein" (Skala 3) Modellunverträglichkeiten aufweisen; in geringerem Umfang ist dies auch für Item Nr. 2 "Schuldgefühle" in beiden Subskalen der Fall; letzteres gilt auch für die Subskala mit Itemumfang fünf (Skala 4). Doch da die Modellunverträglichkeit des Items in der Subskala mit fünf Items nicht zu einer Modellunverträglichkeit der gesamten Subskala führt, ist sie zu tolerieren. Gleichwohl sind die geringeren Modellunverträglichkeiten auf Itemniveau bei der Subskala mit Itemumfang fünf (Skala 4) ein Argument, diese Skala als die relativ optimale anzusehen.

Bei der Interpretation der Korrelation mit den Summenscores von konkurrierenden Depressionsskalen und mit der Globalbeurteilung des Schweregrades ist als methodischer Mangel die fehlende Unabhängigkeit der Erhebung der Beurteilungen (zwischen der zu validierende HAMD-Skala bzw. Subskalen einerseits und den Validierungskriterien MES-, MADRS- und Globalbeurteilungsskala andererseits) in Rechnung zu stellen. Die Korrelationen mit der Schweregradbeurteilung nach MES und nach der Globalbeurteilung weisen die Subskala mit Itemumfang fünf als optimal unter den geprüften HAMD-Skalen bzw. -Subskalen aus; die Korrelation der HAMD-Skalen (17-Items-Version) mit der MADRS-Skala (Summenscore) ist dagegen günstiger als die der HAMD-Subskalen. Insgesamt führt also die fünf Items umfassende Subskala im Vergleich zur 17-Items-Version der HAMD nicht zu einem Validitätsverlust (Korrelation mit anderen Schweregrad-



parametern), sondern zu einem leichten Validitätsgewinn der Schweregradbeurteilung depressiver Syndrome.



## 7. ZUSAMMENFASSUNG

Es wird untersucht, ob die Hamilton-Depressions-Skala (HAMD) ein geeigneter Maßstab zur Beurteilung des Schweregrades des depressiven Syndroms oder von Prägnanztypen des depressiven Syndroms ist. Als Kriterium hierfür wurde die Homogenität und die Stichprobenunabhängigkeit der Struktur der Skala bzw. der Subskalen verwendet. Hierzu wurde die HAMD in zwei verschiedenen Stichproben stationärer Patienten zum Zeitpunkt der Aufnahme bewertet; zusätzlich wurde die HAMD in einer dieser Stichproben drei Wochen nach Beginn der antidepressiven Therapie neuerlich erhoben. Die Analyse der HAMD-Daten erbrachte die folgenden Resultate:

1. In der HAMD (Version mit 17 bzw. 21 Items) kann faktorenanalytisch in allen untersuchten Stichproben nur ein stabiler Faktor extrahiert werden.
2. Dieser stabile Faktor beschreibt in keiner Stichprobe einen Prägnanztyp des depressiven Syndroms; sie tragen auf mindestens einem Viertel der Items jedoch irrelevante Ladungen. Daher sind sie nur unzureichend als "Schweregradfaktoren" interpretierbar. Es empfiehlt sich daher eine Reduktion der HAMD, in der jedenfalls jene Items eliminiert werden, die auf den stabilen Faktoren irrelevante Ladungen tragen.
3. Die Homogenitätsanalysen (nach den Methoden von MOKKEN bzw. RASCH) erwiesen die HAMD in allen Stichproben als unzureichend homogen. Eine Reduktion der HAMD wird dadurch nahegelegt.
4. Bei der Reduktion der HAMD nach der von MOKKEN vorgeschlagenen Methode resultiert eine Skala vom Itemumfang sechs als umfangreichste homogene Subskala; die sechs Items sind: "depressive Stimmung" (Item Nr. 1), "Schuldgefühle" (Item Nr. 2), "Arbeit und sonstige Tätigkeiten" (Item Nr. 7), "Hemmung" (Item Nr. 8), "Erregung" (Item Nr. 9), "Psychische Angst" (Item Nr. 10).



5. Diese Subskala stimmt nur in einem Item nicht mit der von BECH et al. mit anderen Mitteln und in Stichproben anderer Zusammensetzung isolierten Subskala (Itemumfang sechs) überein; anstelle des Items "Agitation (Item Nr. 9)" findet sich in der Subskala von BECH das Item "körperliche Symptome - allgemein (Item Nr. 13)".
6. Als homogenste und valideste Subskala mit möglichst großem Umfang erwies sich in allen untersuchten Stichproben die Skala von Itemumfang fünf, die sowohl in der Subskala von BECH als auch in der hier isolierten Subskala enthalten ist. Die Validität dieser Subskala ist gesichert; die Reliabilität konnte allerdings noch nicht ausreichend geprüft werden.



## 8. LITERATUR

1. BAUMANN, U.: (1976) Methodische Untersuchungen zur HAMILTON-Depression's-Skala Arch. Psychiat. Nervenkr. 222, S. 359-375.
2. BECH, P.: (1981) Rating scales for affective disorders: their validity and consistency, Acta Psychiatrica Scand. Suppl. 295.
3. BECH, P.: (1984) The instrumental use of rating scales of depression, Pharmakopsychiatrie 17, S. 63-75.
4. BECH, P., ALLERUP, P., GRAM, F., REISBY, N., ROSENBERG, R., (1981) The JACOBSEN, O., Hamilton Depression Scale, Acta Psychiat. NAGY, A.: Scand. 63, S. 290-299.
5. BECKMANN, H., SCHMAUSS, M.: (1983) Clinical Investigations into anti-depressivemechanisms. I. Antihistaminic and Cholonytic effects: Amitryptilin versus Promethazine. Arch Psychiat. Nervenkr. 233, S. 59-71.
6. BENKERT, O., HOLSBOER, F., ORENGO, P.: (1981) Zum Wirksamkeitsnachweis von Antidepressiva. In: LECHNER, H., KUGLER, J., FONTANARI, D. (Hrsg.). Depressive Syndrome, Alkoholismus, Amsterdam S. 102-113.
7. BENKERT, O., MAIER, W.: (1981) Biologisch-psychiatrische Forschung und psychiatrische Pharmakotherapie München. Med. Wochenschrift 123, S. 795-799.
8. CATTELL, R. B.: (1973) Personality and Mood by Questionnaire, San Francisco.
9. CINCETTI, V., PRUSOFF, B. A.: (1983) Reliability of Depressions and Associated Clinical Symptoms, Arch. Gen. Psychiat. 40, S. 987-990.
10. CIPS - Collegium Internationale Psychiatricae Sclalarum: (1977) Internationale Skalen für Psychiatrie, Berlin.
11. CLEARY, P., GUY, W.: (1977) Factor analysis of the Hamilton Depression Scale, Drug expt. clin. Res. 1, S. 115-120.
12. COHEN, J.: (1983) The cost of dichotomization, Appl. Psychol. Measurement 7, S. 249-255.
13. DECKWITZ, R., HELMCHEN, H., KOCKOTT, G., MOMBOUR, W.: (1980) Diagnose-schlüssel und Glossar psychiatrischer Krankheiten, 9. Revision, Berlin.
14. ENDICOTT, J., COHEN, J., NEE, J., FLEISS, J., SARANTAKOS, S.: (1981) Hamilton Depression Scale, Arch. Gen. Psychiat. 38, S. 98-103.



15. FEIGHNER, J. R., BRANZER, B., GELENBERG, A. J., GOMEZ, E., KIEV, A., KORLAND, M. L.: (1979) A placebo controlled multicenter trial of limbatril versus its components in symptomatic treatment of depressive illness. *Psychopharmacology* 61, S. 217-225.
16. FISCHER, G.: (1974) Einführung in die Theorie psychologischer Tests, Bern.
17. FISCHER, G., ROPPERT, J.: (1964) Bemerkungen zu einem Verfahren der Transformationsanalyse, *Arch. ges. Psychol.* 116, S. 98-100.
18. GEBHARDT, F.: (1968) Die Ähnlichkeit von Faktorenmatrizen, *Psych. Beiträge* 11, S. 591-599.
19. HAMILTON, M.: (1960) A rating scale for depression, *J. Neurol. neurosurg. Psychiat.* 23, S. 56-62.
20. HAMILTON, M.: (1967) Development of a rating scale for primary depressive illness, *Brit. J. soc. clin. Psychol.* 6, S. 278-296.
21. HEDLUND, I. L., VIEWEG, B. W.: (1979) The Hamilton Rating Scale for Depression. A comprehensive review. *J. Operational Psychiat.* 10, S. 149-165.
22. HEIMANN, H.: (1975) Methodologische Probleme bei der Effizienzprüfung von Psychopharmaka, *Arch. Psychiat. Nervenkr.* 220, S. 255-268.
23. HODDERN, A., BURT, C. G., HOLT, N. F., GADE, J. F. J.: (1965) *Depressive States*, Springfield.
24. JÄGER, A. G., HORMANN, H. J.: (1981) Determination of "g" (allgemeine Intelligenz) und die Bedeutung des Variablenkontextes bei explorativen Faktorenanalysen, *Psych. Beiträge* 23, S. 403-420.
25. KATZ, M. M., KOSLOW, S. W., MAAS, J. W., BOWDEN, C. L., CROUGHAN, J., KOLSI, J.: (1984) Subtypes of depression and response to tricyclic drugs in: RACAGNI, G., PAOLETTI, R., KIELHOLZ, P.: *Clinical Neuropsychopharmacology*, Florence.
26. KEARNS, N. P., CRUICKSHANK, C. A., Mc QUIGAN, R. J., RILEY, S. A., SHAW, S. P., SNAITH, R. P.: (1982) A comparison of depression rating scales. *Brit. J. Psychiatry* 141, S. 45-49.
27. KIELHOLZ, P. (Hrsg.): (1973) *Die larvierte Depression*, Bern.
28. KIELHOLZ, P.: (1971) *Diagnose und Therapie der Depression für den Praktiker*, 3. Auflage, München.



29. LIENERT, G. A.: (1972) Testaufbau und Testanalyse, Meisenheim.
30. LIPPERT, E., SCHNEIDER, P., WAKENHUT, R.: (1978) Die Verwendung der Skalierungsverfahren von MOKKEN und RASCH zur Überprüfung und Revision von Einstellungsschaden. Diagnostica 24, S. 252-275.
31. LOEVINGER, J.: (1947) A systematic approach to the construction and evaluation of tests of ability. Psychol. Monogr. 64, S. 285.
32. MICHAUX, M. H., SUIDELES, A., GARMIZE, C., ROSSI, A. J.: (1969) Depression factors in depressed and in heterogeneous inpatient sample, J. Neurol. Neurosurg. Psychiat. 32, S. 609-613.
33. MICHAUX, M. H., ZWAAG, L. V., KURLAND, A. A.: (1967) Dimensions of manifest depression in newly hospitalized patients, Brit. J. Psychiat. 113, S. 981-986.
34. MÖLLER, H. J., ZERSSEN, D.: (1983) Psychometrische Verfahren II: Standardisierte Beurteilungsverfahren, Nervenarzt 54, S. 1-16.
35. MOKKEN, P.: (1972) Scale analysis, University Press, Den Haag.
36. MONTGOMERY, P., ASBERG, M.: (1979) A new depression scale designed to be sensitive to change, Brit. J. Psychiat. 134, S. 382-389.
37. MOWBRAY, R. M.: (1972) The Hamilton Scale for Depression: A factor analysis, Psychol. Med. 2, S. 272-280.
38. RASCH, G.: (1960) Probability models for some intelligence and attainment tests, Kopenhagen.
39. RHOADES, M., OVERALL, J. E.: (1983) The Hamilton-Depression-Scale: factor scoring and probit-scoring, Psychopharmacology Bulletin 19, S. 121-125.
40. PAWLIK, K.: (1968) Dimensionen des Verhaltens, Huber-Verlag, Bern.
41. PRUSSOF, B., WEISSMAN, M. M.: (1981) Pharmacologic treatment of anxiety in depressed outpatients. In: KLEIN, D. F., RABKIN, J. G. (eds) Anxiety: New Research and changing concepts, New York.
42. RAFAELSEN, O. J., BECH, P., BOWLIG, T. G., KRAMP, P., JERRIS, A.: (1980) The Bech Rafaelsen Combined Rating Scale for Mania and Melancholia. In: ACHTE, K., AALBERG, V., LÖNNQUIST, J. (eds): Psychopathology of depression, Acta Psychiat. Fennica Supplementum (1980).



43. REVENSTORF, D.: (1976) Lehrbuch der Faktorenanalyse, Stuttgart.
44. SCHATZBERG, A., COLE, J. O.: (1978) Benzodiazepine in depressive disorders. Arch. Gen. Psychiatry 35, S. 1359-1365.
45. SCHNEEWIND, K. A., CATTELL, R. B.: (1970) Zum Problem der Faktorenidentifikation, Psychol. Beiträge 12, S. 214-226.
46. SCHNEIDER, K.: (1955) Klinische Psychopathologie, 4. Auflage, Stuttgart.
47. SMOUSSE, P. E., FEINBERG, M., CARROLL, B. J., PARK, M. H., RAWSON, S.G.: (1981) The Carroll Rating Scale for Depression II, Factor-Analysis of Feature Profiles, Brit. J. Psychiat. 138, S. 201-204.
48. SPITZER, R. L., ENDICOTT, J., ROBINS, E.: (1978) Research Diagnostic Criteria: Rational and Reliability, Arch. Gen. Psychiat. 35, S. 773-785.
49. WEISSMANN, M. N., KLERMAN, L., PRUSSOFF, B. A., SHOLOMSKAS, D., PADIAN, N.: (1981) Depressed outpatients, Arch. Gen. Psychiat. 38, S. 51-55.
50. WEITBRECHT, H. J.: (1972) Depressive und manische endogene Psychosen in: KISKER, K. P., MEYER, J. E., MÜLLER, M., STRÖMGREN, E., Psychiatrie der Gegenwart Band II, Teil 1, Berlin.
51. WELNER, J.: (1972) Eine internationale multizentrische Doppelblindprüfung eines neuen Antidepressivums. In: KIELHOLZ, P. (Hrsg): Depressive Zustände Huber-Verlag, Bern.
52. WING, J. K., COOPER, J. E., SARTORIUS, N.: (1978) Present State Examination (Deutsche Bearbeitung), Weinheim.



## Lebenslauf

Geburtsdatum: 13.01.1949

Geburtsort: Hof/Saale

1955 - 1959 Besuch der Volksschule in Hof

1959 - 1968 Besuch des Schillergymnasiums in Hof

1968 Abitur

1968 - 1974 Studium der Mathematik und Philosophie an der Ludwig-Maximilians-Universität München

1974 Abschluß als Diplommathematiker

1974 - 1978 Beschäftigung als geprüfte wissenschaftliche Hilfskraft am Mathematischen Institut bzw. am Institut für Wissenschaftstheorie und Statistik der Universität München

1974 - 1980 Studium der Humanmedizin an der Ludwig-Maximilians-Universität München

1980 Approbation als Arzt

seit Juli 1981 Beschäftigung als wissenschaftlicher Angestellter (Arzt) an der Psychiatrischen Klinik der Johannes-Gutenberg-Universität Mainz









